

DOCUMENT RESUME

ED 178 967

CS 502 682

TITLE Status Report on Speech Research: A Report on the Status and Progress of Studies on the Nature of Speech, Instrumentation for Its Investigation, and Practical Application, April 1-June 30, 1979. Status Report No. 58.

INSTITUTION Haskins Labs., New Haven, Conn.

PUB DATE 79

NOTE 202p.

EDRS PRICE MF01/PC09 Plus Postage.

DESCRIPTORS *Articulation (Speech); Beginning Reading; Japanese; Language Patterns; Language Skills; Phonetics; Physiology; Reading Skills; *Research; Semantics; Speech; *Speech Communication; *Speech Skills; Vowels

IDENTIFIERS *Oral Language

ABSTRACT

This report is one in a regular series on the status and progress of speech research. The 11 manuscript topics are: towards a theoretical reassessment of the role of proprioception in the perception and control of human movement, interarticulator programming in stop production, the beginnings of speech, proofreading errors on the word "the," the speech code and learning to read, some experiments on the sound of silence in phonetic perception, phonological coding in beginning reading, laryngeal adjustments during Japanese fricative and devoiced vowel production, on the coordination of two-handed movements, the contribution of natural durations to speech synthesized by F0V0 rules, and a review of "The Skills of the Plodder." (FL)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED178967

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

SR-58 (1979)

Status Report on

SPEECH RESEARCH

A Report on
the Status and Progress of Studies on
the Nature of Speech, Instrumentation
for its Investigation, and Practical
Applications

1 April - 30 June 1979

Haskins Laboratories
270 Crown Street
New Haven, Conn. 06510

Distribution of this document is unlimited.

(This document contains no information not freely available to the general public. Haskins Laboratories distributes it primarily for library use. Copies are available from the National Technical Information Service or the ERIC Document Reproduction Service. See the Appendix for order numbers of previous Status Reports.)

CS502682

ACKNOWLEDGMENTS

The research reported here was made possible in part by support from the following sources:

National Institute of Child Health and Human Development
Grant HD-01994

National Institute of Child Health and Human Development
Contract NO1-HD-1-2420

National Institutes of Health
Biomedical Research Support Grant RR-05596

National Science Foundation
Grant BNS76-82023

National Science Foundation
Grant MCS76-81034

National Institute of Neurological and Communicative
Disorders and Stroke
Grant NS13870
Grant NS13617

National Institute of Arthritis, Metabolism, and
Digestive Diseases
Grant AM25814

HASKINS LABORATORIES

Personnel in Speech Research

Alvin M. Liberman,* President and Research Director
Franklin S. Cooper,* Associate Research Director
Patrick W. Nye, Associate Research Director
Raymond C. Huey, Treasurer
Alice Dadourian, Secretary

Investigators

Arthur S. Abramson*
Thomas Baer
Fredericka Bell-Berti+
Catherine Best+
Gloria J. Borden*
Guy Carden*
Robert Crowder*
William Ewan*
Carol A. Fowler*
Jane H. Gaitenby
Thomas J. Gay*
Katherine S. Harris*
Alice Healy*
David Isenberg+
James J. Jenkins³
Leonard Katz⁴
Scott Kelso
Andrea G. Levitt*
Isabelle Y. Liberman*
Leigh Lisker*
Anders Lofqvist²
Virginia Mann+
Charles Marshall
Ignatius G. Mattingly*
Nancy McGarr*
Lawrence J. Raphael*
Bruno H. Repp
Philip E. Rubin
Donald P. Shankweiler*
Winifred Strange³
Michael Studdert-Kennedy*
Michael T. Turvey*
Robert Verbrugge*
Hirohide Yoshioka¹

Technical and Support Staff

Eric L. Andreasson
Elizabeth P. Clark
Donald Hailey
Terry Halwes
Sabina D. Koroluk
Agnes M. McKeon
Nancy R. O'Brien
William P. Scully
Richard S. Sharkany
Leonard Szubowicz
Edward R. Wiley
David Zeichner

Students*

David Dechovitz
Laurel Dent
Laurie Feldman
Hollis Fitch
Carole E. Gelfer
David Goodman
Janette Henderson
Charles Hoequist
Kenneth Holt
Robert Katz
Morey J. Kitzman
Peter Kugler
Roland Mandler
Suzi Pollack
Patti Jo Price
Brad Rakerd
Arnold Shapiro
Louis G. Tassinari
Janet Titchener
Emily Tobey-Cullen
Betty Tuller
N. S. Viswanath
Douglas Whalen

*Part-time

¹Visiting from University of Tokyo, Japan

²Visiting from Lund University, Sweden

³Visiting from University of Minnesota

⁴On sabbatical leave at the University of Sussex, U.K.

+NIH Research Fellow

CONTENTS

<u>I. Manuscripts and Extended Reports</u>	
Towards a Theoretical Reassessment of the Role of Proprioception in the Perception and Control of Human Movement-- J. A. Scott Kelso, Kenneth G. Holt and Adrian E. Flatt	 1
Interarticulator Programming in Stop Production--Anders Lofqvist	 13
The Beginnings of Speech--Michael Studdert-Kennedy	 35
Proofreading Errors on the Word <u>The</u> : New Evidence on Reading Units--Alice F. Healy	 65
The Speech Code and Learning to Read--Donald Shankweiler, Isabelle Y. Liberman, Leonard S. Mark, Carol A. Fowler, and F. William Fischer	 83
Some Experiments on the Sound of Silence in Phonetic Perception-- Michael F. Dorman, Lawrence J. Raphael and Alvin M. Liberman	 105
Phonological Coding in Beginning Reading--Carol A. Fowler	 139
Laryngeal Adjustments during Japanese Fricative and Devoiced Vowel Production--Hirohide Yoshioka	 147
On the Coordination of Two-Handed Movements--J. A. Scott Kelso, Dan L. Southard and David Goodman	 161
The Contribution of Natural Durations to Speech Synthesized by FOVE Rules--Frances Ingemann	 177
A Review of "The Skills of the Plodder"--Ignatius G. Mattingly	 185
<u>II. Publications</u>	 191
<u>III. Appendix: DDC and ERIC numbers (SR-21/22 - SR-58)</u>	 193

I. MANUSCRIPTS AND EXTENDED REPORTS

TOWARDS A THEORETICAL REASSESSMENT OF THE ROLE OF PROPRIOCEPTION IN THE PERCEPTION AND CONTROL OF HUMAN MOVEMENT*

J. A. Scott Kelso,+ Kenneth G. Holt,+ and Adrian E. Flatt++

Abstract. The theoretical role of proprioception in the perception and control of human movement is elusive because of the obvious inability to manipulate the various receptive systems experimentally. Individuals who have had the metacarpophalangeal joint and joint capsule removed and replaced with silastic inserts afford a unique opportunity to evaluate a principal source of proprioception, namely, slowly adapting joint afferents. In a set of experiments we show that such individuals show no deficits in finger localization following joint replacement. We take this and other complementary findings as a basis for proposing a dynamic rather than kinematic account of movement production. In addition, we provide a reconceptualization of the function of proprioceptive information in the central nervous system. Our arguments focus on proprioceptive inputs as tuning or modulating interneuronal pools rather than providing dimension-specific information to the brain as is commonly assumed.

An important limitation for those of us who seek to understand the control of human movement is that we are, by necessity, confined to observations about motor output upon which to infer the nature of the underlying processes involved. It is always difficult to discern which aspects of the motor output represent central control and which components reflect peripheral contributions. A major tack on approximating the peripheral informational support for human movement is to use techniques designed to interrupt or disrupt afferent function. Unfortunately, procedures that have been adopted thus far that attempt to interfere with the flow of kinesthetic information to the central nervous system via peripheral nerve blocks are rife with problems (Kelso, Stelmach, & Wanamaker, 1974).

* A preliminary version of this paper was presented at the Psychonomic Society Meetings, San Antonio, Texas, November 1978.

+Also University of Connecticut, Storrs.

++Department of Surgery, Norwalk Hospital, Norwalk, Connecticut.

Acknowledgment: This research was supported by Grants AM-25814 and NS13617 from the National Institutes of Health. We would like to thank Franklin S. Cooper and Michael T. Turvey for their helpful remarks on an earlier version.

What is required is a preparation that selectively eliminates an important source of kinesthetic input without significantly impairing peripheral motor structures. Surgical operations carried out in humans that involve the replacement of joints provide a potential means for deriving important inferences on the role of kinesthesia in movement perception and control. Common to such procedures is the fact that the joint capsule--which purportedly accommodates receptors for position and movement (Skoglund, 1956)--is completely removed and the joint surfaces replaced. The patient therefore provides a unique opportunity to examine motor performance in the absence of the capsular component of peripheral receptor mechanisms. This is of particular significance, for recent theoretical papers and many critical reviews (e.g., Mountcastle, 1968; Roland, 1978) refer invariably to joint receptors as detectors of joint angle, and even as crucial to motor timing (Adams, 1977). Neither the current physiological data on joint receptors nor the behavioral data that we shall present support such a proposition. In contrast, our findings indicate that joint receptors are not necessary for detecting limb position. Moreover, they are extremely unlikely candidates for primary status in the temporal control of movement. We take advantage of our findings to elaborate upon a new style of control--initially promoted by Soviet theorists and developed by Turvey and others (Bernstein, 1967; Fitch & Turvey, 1977; Greene, 1972; Turvey, 1977)--that fits our general perspective on the nature of coordinated movement (Kelso, Southard, & Goodman, 1979a, 1979b).

The work of Skoglund (1956) and Boyd (1953) is typically regarded as a demonstration that joint afferent discharge is angular specific. Thus single neurons from slowly adapting receptors in the capsule of the cat knee-joint were shown to fire maximally at particular joint angles and with a sensitive range of 15 to 30 degrees. Unfortunately, recent and expansive data fail to confirm the early findings that joint afferents discharge at intermediate angles although supporting the view that much more activity is seen at the very extremes of flexion and extension (Burgess & Clark, 1969; Grigg & Greenspan, 1977; Lynn, 1975). In fact, when the popliteus muscle, which is located posterior to the knee-joint, is carefully removed, the midrange response is eliminated (Clark, 1975). Furthermore, the small number of midrange fibers found are strongly sensitive to succinylcholine chloride, a drug that is selectively responsive to muscle receptors. In contrast, no such sensitivity is observed in joint receptors that fire at the end of the movement range (Clark & Burgess, 1975). Obviously it would be useful to corroborate these new neurophysiological data with information from humans who have lost joint capsular afferents. In fact, some work has already been done in this regard. Grigg, Finerman, and Riley (1973), for example, performed a number of psychophysical tests on patients who had undergone total hip replacement. Their results revealed that loss of position sense was restricted, but only slightly so, to passive movements. Active movements showed no such deficit. It seems possible, however, that this result may be confounded with the fact that the hip joint is intrinsically involved in locomotory activities which, if animal evidence is a guide, do not require ongoing kinesthetic information (Grillner, 1975). Thus we might expect to see considerable differences in kinesthetic sensitivity between hip and finger joints, for example. Indeed, a simple comprehensive statement about the general properties of joint afferents across different joints has proved somewhat elusive. Although observations of knee and elbow joints converge for cats and primates in failing to show midrange responses, evidence from

costovertebral (Godwin-Austin, 1969) and temporomandibular joints (Thilander, 1961) indicates the presence of full range receptors.

In the present experiments we examined the accuracy of movement reproduction of the index finger following complete surgical removal of the metacarpophalangeal (MP) joints in the hands of patients suffering from rheumatoid arthritis. In all cases the inserted "prosthesis" was one developed by Swanson (1972) and made of silastic rubber. In essence the device is not so much an articulated prosthesis as an implant designed to hold apart the two bone surfaces of the metacarpal and the proximal phalanx. Most patients had all four MP joints replaced, and all patients had the MP joint of either the right or left index finger removed. The movements allowed by the positioning device were flexion and extension of the index finger about the MP joint. The distal end of the finger was fitted with a plastic collar that slipped into an open-ended cylindrical support. The support revolved around the MP joint and prevented movement of the distal joints of the finger. Attached to the end of the support was a pointer that moved over a protractor graduated in degrees. The device was also equipped with padded adjustable clamps with which to secure the patient's wrist, hand and remaining fingers and thumb during the movement. Only the preferred hand was placed in the device while the other rested on the patient's lap. Vision of the hand was obscured by an aluminum screen. Procedures closely followed previous work (Kelso, 1977). In a preliminary study patients (n=5) and normal subjects (n=12) performed 12 preselected and 12 constrained movements into each of three movement sectors defined initially by the experimenter. Thus, for preselected movements, instructions were to "select" a short, medium, or long movement of the finger and then, following a 2-sec interval, to "move" to the desired position. In this case, therefore, subjects were free to choose their own desired movements with the only restriction that they disperse their selections within a given sector as much as possible. For constrained movements the commands were "ready" followed by "move" and the patient moved until he or she located a mechanical stop defining the movement. Thus subjects made constrained, exploratory movements since no prior selection was possible. Patients in both conditions reproduced the criterion movement with the stop removed following their passive return to the starting position, which remained constant throughout.

A main feature of our data was that there were minimal differences between normal subjects and joint replacement patients. On preselected movements the mean reproduction error of normal subjects was 2.98 degrees (1 degrees = 2 mm measured at the tip of the index finger) compared to 3.13 degrees for the joint replacement group. Although errors on constrained movements were slightly higher overall, the remarkable result was that the removal of joint afferent information had no effect whatsoever (means = 4.44 degrees and 3.97 degrees for normals and joint replacement patients, respectively). This finding was in sharp contrast to a situation where normal subjects (n=12) performed under conditions where joint and cutaneous information were eliminated via the application of a child's sphygmomanometer (blood pressure cuff) at the wrist. This technique has the advantage of preserving muscle function in finger flexors and extensors since these muscles lie high in the forearm above the cuff (Goodwin, McCloskey, & Matthews, 1972; Kelso, 1977; Merton, 1964). Although preselected performance was hardly affected (mean = 3.34 degrees), there were considerable deleterious effects under

constrained, exploratory conditions (mean = 13.34 degrees). Indeed, phenomenological reports revealed that wrist cuff subjects could not perceive the locus of the mechanical stop when performing constrained movements. This was not the case for joint replacement patients.

While these data are highly suggestive that joint afferent information is not crucial for the perception and control of movement, we must emphasize that patients in our initial experiment varied in the extent of the post-operative recovery period from six weeks in one case to over a year in another. An examination of the individual data, however, did not reveal any sizable systematic differences among patients as a function of the post-operative period. Nevertheless, it would clearly be more satisfactory to collect data from patients as soon after the operation as possible.

The follow-up experiments were on 13 patients who were examined during a period from two days to four weeks following total MP joint arthroplasty. On some occasions pre-tests were given using the same experimental paradigm as discussed above. However, we do not consider differences between pre- and post-test reliable because of a number of potentially confounding factors: for example, stiffness of the joints prior to operation, anxiety, etc. In fact the direction of the difference, if one existed, was in favor of post operation performance.

The basic experimental procedure in this study (termed Experiment 1) was, with one exception, identical to that employed in our preliminary work. Patients performed 12 preselected, constrained and passive trials into one of three movement sectors (see the caption of Figure 1 for details). Absolute (unsigned), constant (signed), and variable (the standard deviation around the mean constant error) errors were collapsed across sectors and analyzed in a 3 (movement conditions) x 3 (movement sector) analysis of variance. The main effect of movement conditions was significant for absolute and variable error only, $F(2, 24) = 12.30$, $p < .001$, and $F(2, 24) = 2.93$, $p < .05$, respectively. For absolute error, preselected movements were superior to passive and constrained, which were not different from each other (see Figure 1). A similar pattern of results obtained for variable error. In this case preselected was superior to passive but not constrained conditions, although the latter two were not different from each other.

Neither the sectors' main effect nor the interaction of sectors and conditions was significant for any of the dependent variables. The superiority of preselection over constrained and passive conditions shown in Figure 1 replicates much of our previous work and has been discussed in detail elsewhere (Jones, 1974; Kelso, 1977; Kelso & Stelmach, 1976). But the most interesting finding for the present discussion is the level of error in constrained and passive conditions. It is quite obvious that the patients in this study compare favorably with their counterparts in our preliminary experiment; more importantly, they perform within normal ranges. This is a fascinating finding, particularly in light of the classical role given by most physiologists and psychologists to joint receptors in the perception of movement and position (Mountcastle, 1968; Roland, 1978; Skoglund, 1956; Somjen, 1972). That is, "classical" conceptions of kinesthesia are built upon the angular specificity viewpoint--a contrary notion to very recent physiological work and obviously at variance with our data.

But what are the alternatives to joint receptors? Whether tactile information is sufficient to account for the performance of joint replacement patients is open to question. Goldscheider's (1889) work in which the skin was anesthetized via an AC electric current revealed no disturbing effects on movement perception. More recently, however, the Swedish surgeon Moberg (1972), in a unique patient, has shown that although joint receptor information was unavailable, perception of passive motion and position was preserved with only skin receptors in function. Another alternative is that cutaneous inputs facilitate access to the central nervous system by muscle receptors. If this is the case, a strong argument could be generated for the role of muscle receptors in the conscious appreciation of movement—a stance that is receiving increasing support (Matthews, 1977).

We should note that patients in previous experiments had several sources of information available to them that may have assisted accurate movement production. Patients knew, for example, that the starting position of the finger remained the same throughout testing. Thus they could use other information—such as duration or velocity—as a means for arriving at the correct final position. We examined this proposition by considering performance under conditions where the starting position changed for the reproduction movement, thereby disrupting either the amplitude moved or the final end position reached. Under one condition the patient was asked to produce the final position, while another condition required the patient to reproduce the same amplitude or distance (see Figure 2 for details). For absolute and constant error there was a significant interaction between movement conditions and starting position, $F(1, 12) = 7.76, p < .02$ and $F(1, 12) = 11.27, p < .01$, respectively. It is clear that location is superior overall to amplitude and that the effect is magnified at the extreme starting position. Interestingly, amplitude performance is biased in the direction of the final position presented. Thus while location performance is hardly affected by changes in starting position, amplitude performance appears to reflect a bias to reproduce location. This finding suggests rather strongly that location is the important "code"; even though instructed to reproduce amplitude, the motor system appears to be optimally organized for achieving final position. The latter, we emphasize, does not crucially depend upon slowly adapting joint receptors. Indeed, in earlier work on normal subjects a very similar finding was obtained between amplitude and location when both joint and cutaneous inputs were removed (Kelso, 1977).

One way of interpreting these data is that there is a location code based on information provided by some type of peripheral receptor or set of receptors. Reproduction of location may then be viewed as a matching of receptor inputs to the stored referent or spatial code. Reproduction of distance, however, is much more difficult in that the change in starting position requires an additional subtractive process relative to the spatial code. Thus to reproduce accurately, a new spatial code must somehow be derived to take into account the change in starting position (Stelmach & McCracken, 1978).

An alternative explanation, and one that has gained status in recent papers (Bahill & Stark, 1979; Bizzi, Dev, Morasso, & Polit, 1978; Kelso, 1977), takes advantage of the natural physical properties such as damping, stiffness, and inertial resistance that are inherent in neuromuscular control

systems. Typically, muscle-joint linkages are viewed as homeomorphic vibratory systems, the most specific example being a mass-spring (Asatryan & Fel'dman, 1965; Fel'dman, 1966). Our findings may be interpreted as displaying an important characteristic of a mass-spring system, namely that of equifinality (von Bertalanffy, 1973). That is, despite changes in initial conditions (displacement of a limb to a new starting position, mechanical perturbations), a mass-spring system will always reach an invariant final position or equilibrium point, determined only by the parameter specifications. For example, Polit and Bizzi (1978) in their recent work, trained monkeys to point with an unseen arm to target lights. At random intervals and prior to pointing, a torque motor displaced the arm further away from, closer to, or even beyond the target. In spite of such alterations of kinesthetic input the final position was always reached. These data suggest that final position is determined via the specification of stiffness and damping parameters that establish an equilibrium point between opposing pairs of muscles. That kinesthetic information is not crucial to this type of mechanism is revealed by identical results in animals who have undergone bilateral dorsal rhizotomy.

An argument can be made, therefore, that the neuromuscular organization underlying achievement of location has the features of a vibratory system. Note that the two viewpoints discussed here differ considerably in perspective. The former argues that the kinematic details of movements are represented in a spatial code. Thus location as the endpoint of a movement may be described in reference to some internal coordinate system. While this may be a legitimate description, it refers to kinematics and not dynamics. The point should be clear when it is realized that it is the dynamics (e.g., force, viscosity, etc.) that determine the movement kinematics. From a dynamic perspective then, terminal location is equated with the steady state of a system and is determined only by the parameters selected. Nowhere is there a need to represent kinematic details: It is in the nature of a vibratory system to achieve equilibrium. While the present experiments cannot entirely differentiate these alternatives, the parsimony of the dynamic description is appealing. The vibratory system viewpoint clearly negates reliable reproduction of distance (a kinematic detail) from variable initial conditions. Furthermore, that accurate achievement of final position can obtain in the absence of slowly adapting joint afferents muddies the common view that angular specific receptors contribute to the development of a spatial code.

Finally, it behooves us to consider briefly the theoretical role that joint receptors may play in the control of movement. One possibility arises out of Grigg's (1976) work showing that a sizable proportion of afferents in cat medial nerve fire as a function of the degree of torque developed at a fixed joint position. This finding suggests that muscular contractions activate joint neurons and that joint afferents can function as load detectors. But another, more intriguing notion with potentially broad theoretical consequences may be found in a diversion away from traditional views of peripheral mechanoreceptors. Such receptors have typically been regarded as contributing--or not contributing--specific types of kinematic information (e.g., position, rate, acceleration) to higher brain centers for use in control and termination of movements. Suppose, however, that peripheral receptor information is not dimension-specific; rather it serves merely to bias interneuronal pools in the spinal cord so as to lower the threshold at

which signals may be generated to the musculature. Thus the function of mechanoreceptors is simply to "tune" the interneuronal pool so that central command pulses may have an optimal facilitatory effect on the muscles served by that pool. The research of Aizerman and his colleagues (Aizerman & Andreeva, 1968; Chernov, 1968; Litvintsev, 1972) has provided evidence for this viewpoint with reference to muscle spindle function in such activities as postural adjustment, pain avoidance, and precision aiming. For example, if a person in a relaxed standing position is pushed in the back, the spindles in the gastrocnemius and hamstring muscle groups will be stretched. An undifferentiated supraspinal command pulse results in the activation of only those muscles whose spindle inputs define the background state of the interneuronal pool. Consequently, selective activation of the stretched muscles automatically gives rise to forces that preserve vertical posture. It seems imminently possible that the control system may also use cutaneous and joint inputs to serve similar "tuning" functions. In fact, when we realize that human stretch reflex function is virtually eliminated when joint and cutaneous information is removed (Maraden, Merton, & Morton, 1972), this hypothesis gains respectability. The obvious beauty of such a system is that the brain does not have to select which muscles to contract; rather, muscles are activated by virtue of the dynamic state of the interneuronal pools.

REFERENCES

- Adams, J. A. Feedback theory of how joint receptors regulate the timing and position of a limb. Psychological Review, 1977, 84, 504-523.
- Aizerman, M. A., & Andreeva, E. A. Simple search mechanism for control of skeletal muscles. Automation and Remote Control, 1968, 29, 452-463.
- Asatryan, D. G., & Fel'dman, A. G. Functional tuning of the nervous system with control of movement or maintenance of a steady posture -I. Mechanographic analysis of the work of the joint on execution of a postural task. Biophysics, 1965, 10, 925-935.
- Bahill, A. T., & Stark, L. The trajectories of saccadic eye movements. Scientific American, 1979, 240, 108-117.
- Bernstein, N. The co-ordination and regulation of movements. London: Pergamon Press, 1967.
- Bizzi, E., Dev, P., Morasso, P., & Polit, A. Effect of load disturbances during centrally initiated movements. Journal of Neurophysiology, 1978, 41, 542-555.
- Boyd, I. A., & Roberts, T. M. Proprioceptive discharge from stretch-receptors in the knee joint of the cat. Journal of Physiology, 1953, 122, 38-58.
- Burgess, P. R., & Clark, F. J. Characteristics of knee joint receptors in the cat. Journal of Physiology, 1969, 203, 317-335.
- Chernov, V. I. Control over single muscles or a pair of muscle antagonists under conditions of precision search. Automation and Remote Control, 1968, 29, 1090-1101.
- Clark, F. J. Information signalled by sensory fibres in the medial articular nerve. Journal of Neurophysiology, 1975, 38, 1464-1472.
- Clark, F. J., & Burgess, P. R. Slowly adapting receptors in the cat knee joint: Can they signal joint angle? Journal of Neurophysiology, 1975, 38, 1448-1463.
- Fel'dman, A. G. Functional tuning of the nervous system during control of movement or maintenance of a steady posture-III. Mechanographic analysis of the execution by man of the simplest motor tasks. Biophysics, 1966,

11, 765-775.

- Fitch, M. L., & Turvey, M. T. On the control of activity: Some remarks from an ecological point of view. In D. M. Landers & R. W. Christina (Eds.), Psychology of motor behavior and sport. Champaign, Ill.: Human Kinetics Publishers, 1977.
- Godwin-Austin, R. B. The mechanoreceptors of the costo-vertebral joints. Journal of Physiology, 1969, 202, 737-753.
- Goldscheider, A. Untersuchen ueber den Muskelsinn. Archives of Anatomy and Physiology, 1889, 3, 369-502.
- Goodwin, G. W., McCloskey, D. I., & Matthews, P. B. C. The contribution of muscle afferents to kinaesthesia shown by induced illusions of movement by the effects of paralyzing joint afferents. Brain, 1972, 95, 705-748.
- Greene, P. H. Problems of organization of motor systems. In H. Rosen & F. Snell (Eds.), Progress in theoretical biology (Vol. 2). New York: Academic Press, 1972.
- Grigg, P. Response of joint afferent neurons in cat medial articular nerve to active and passive movement of the knee. Brain Research, 1976, 118, 482-485.
- Grigg, P. & Greenspan, B. J. Response of primate joint afferent neurons to mechanical stimulation of knee joint. Journal of Neurophysiology, 1977, 40, 1-8.
- Grigg, P., Finerman, G. A., & Riley, L. H. Joint position sense after total hip replacement. Journal of Bone and Joint Surgery, 1973, 55, 1016-1025.
- Grillner, S. Locomotion in vertebrates: Central mechanisms and reflex interaction. Physiological Reviews, 1975, 55, 247-304.
- Jones, B. Is proprioception important for skilled performance? Journal of Motor Behavior, 1974, 6, 33-45.
- Kelso, J. A. S. Motor control mechanisms underlying human movement reproduction. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 529-543.
- Kelso, J. A. S., Southard, D. L., & Goodman, D. On the nature of human interlimb coordination. Science, 1979, 203, 1029-1031. (a)
- Kelso, J. A. S., Southard, D. L., & Goodman, D. On the coordination of two-handed movements. Journal of Experimental Psychology: Human Perception and Performance, 1979, 5, 229-238. (b)
- Kelso, J. A. S., & Stelmach, G. E. Central and peripheral mechanisms in motor control. In G. E. Stelmach (Ed.), Motor control: Issues and trends. New York: Academic Press, 1976.
- Kelso, J. A. S., Stelmach, G. E., & Wanamaker, W. M. Behavioral and neurological parameters of the nerve compression block. Journal of Motor Behavior, 1974, 6, 179-190.
- Kelso, J. A. S., Wallace, S. A., Stelmach, G. E., & Weitz, G. A. Sensory and motor impairment in the nerve compression block. Quarterly Journal of Experimental Psychology, 1975, 27, 141-147.
- Litvintsev, A. I. Vertical posture control mechanisms in man. Automation and Remote Control, 1972, 33, 590-600.
- Lynn, B. Somatosensory receptors and their central nervous system connections. Annual Reviews of Physiology, 1975, 105-127.
- Marsden, C. D., Merton, P. A., & Morton, H. B. Servo action in human voluntary movement. Nature, 1972, 238, 140-143.
- Matthews, P. B. C. Muscle afferents and kinaesthesia. British Medical Bulletin, 1977, 33(2), 137-142.
- Merton, P. A. Human position sense and sense of effort. Homeostasis and

- feedback mechanisms. 18th Symposium of Society of Experimental Biology. 1964, 18, 387-400.
- Moore, E. Fingers were made before forks. The Hand, 1972, 4, 201-206.
- Mountcastle, V. B. Medical physiology (Vol. 2). St. Louis: Mosby, 1968.
- Polit, A., & Bizzi, E. Processes controlling arm movements in monkeys. Science, 1978, 201, 1235-1237.
- Roland, P. E. Sensory feedback to the cerebral cortex during voluntary movement in man. Behavioral and Brain Sciences, 1978, 1, 129-171.
- Roy, E. A., & Diewert, G. L. Encoding of kinesthetic extent information. Perception and Psychophysics, 1975, 17, 559-564.
- Skoglund, S. Anatomical and physiological studies of knee joint innervation in the cat. Acta Physiologica Scandinavica, 1956, 124, 1-99.
- Somjen, G. Sensory coding in the mammalian nervous system. New York: Appleton-Century-Crofts, 1972.
- Steinbach, G. E., & McCracken, H. D. Storage codes for movement information. In J. Requin (Ed.), Attention and performance VII. Hillsdale, N.J.: Erlbaum, 1978, 515-534.
- Swanson, A. B. Disabling arthritis at the base of the thumb. Journal of Bone and Joint Surgery, 1972, 54-A, 456-471.
- Thilander, B. Innervation of the temporo-mandibular joint capsule in man. Transactions of the Royal Schools of Dentistry (Stockholm), 1961, 7, 49-63.
- Turvey, M. T. Preliminaries to a theory of action with reference to vision. In R. Shaw & J. Bransford (Eds.), Perceiving, acting, and knowing. Hillsdale, N.J.: Erlbaum, 1977.
- von Bertalanffy, L. General systems theory. London: Penguin University Books, 1973.

Figure Captions

Figure 1. Mean absolute error in degrees for joint replacement patients as a function of movement extent. Patients performed 12 preselected and 12 constrained movements whose order was randomly defined. The starting position of the finger on all trials was 20 degrees flexion and the maximum movement seldom exceeded 65 degrees flexion. Patients were instructed to distribute their selections as much as possible within a sector. Preselected movements always came first in the series and constrained movements were yoked to them in order to make an analysis of errors possible. In addition, a passive condition was included in which, following the verbal command "ready," the patient was moved passively to a stop and then returned to the starting position. In all three movement conditions, patients reproduced actively. Although velocities were not measured, the movements in all cases were relatively slow, with an approximate range of 20 to 30 degrees per sec. Time at the endpoint of the movement was held constant at 2 sec.

Figure 2. Mean absolute (unsigned) and constant (signed) error for joint replacement patients as a function of starting position. Under one condition (location) the patient was asked to reproduce final position, while another condition required the patient to reproduce the same amplitude or distance. The same patients participated in this study as in Experiment 1. The criterion movement was presented from a starting position of 20 degrees flexion and was either 35, 45, or 50 degrees flexion, the latter being randomly defined. Patients moved actively to mechanical stops that specified these movements and then were returned to a starting position that was either 5 degrees (SP_1) or 15 degrees (SP_2) beyond the original starting position (i.e., in 15 or 5 degrees flexion). They then reproduced either the final position or the amplitude of movement. Patients performed 12 trials on each condition, with order of position counterbalanced. There were, therefore, two trials on each criterion movement--starting position combination, which were collapsed for inspection of mean absolute and constant errors.

EXPERIMENT I

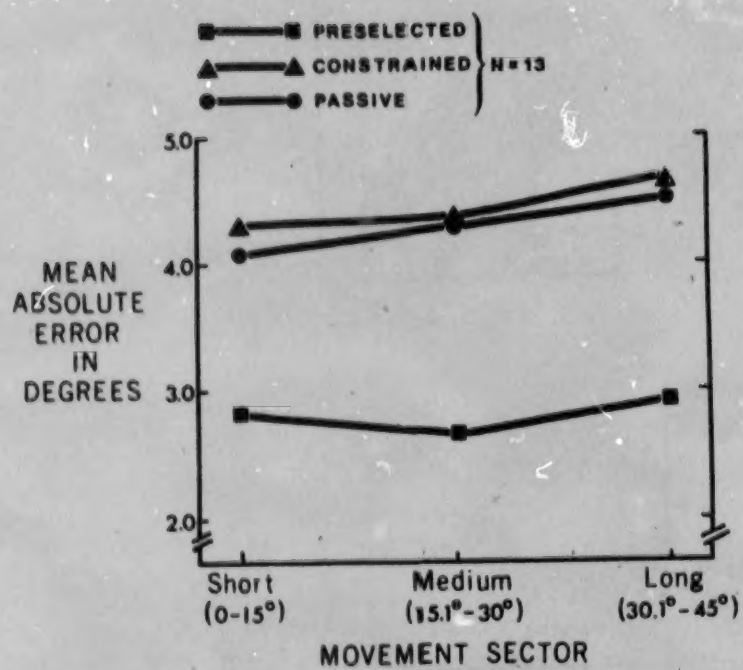


Figure 1.

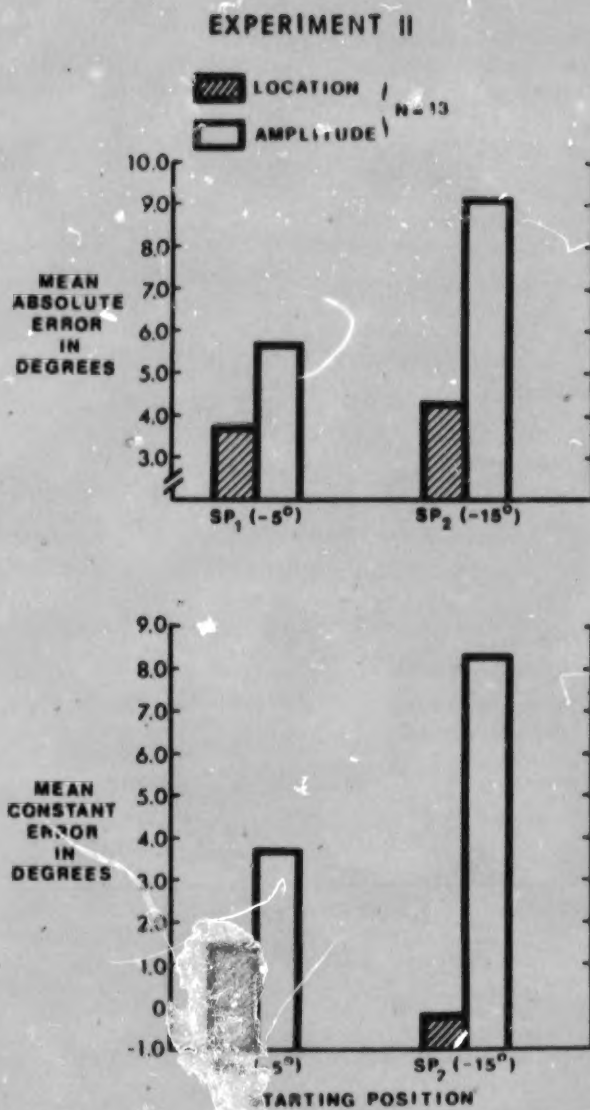


Figure 2.

(This figure accompanies the preceding paper.)

INTERARTICULATOR PROGRAMMING IN STOP PRODUCTION

Anders Lofqvist*

Abstract. The problem of speech motor control has usually been seen as one of accommodating in space and time the articulatory demands for successive units, segments or syllables in the speech chain. Models for speech motor control thus rarely have any intrasegmental temporal domain, but such a domain is necessary for certain classes of speech sounds. The present paper discusses one such instance in the production of Swedish stops.

Voiceless obstruent production requires precise temporal control and coordination of several articulatory systems, and here we examine the coordination of laryngeal and oral articulations in stop production using the transillumination technique and aerodynamic records. The main difference between aspirated and unaspirated stops seems to be one of interarticulator timing, and timing also appears to be the way in which the articulatory system solves the problem of controlling glottal opening at release in aspirated stops. The results are discussed in relation to stop production in general, and some basic characteristics of laryngeal articulatory gestures are outlined as well as some implications for theories of speech motor control.

INTRODUCTION

In a recent paper, Lubker, McAllister, and Lindblom (1977) discuss the notion of interarticulator programming in speech, i.e., the temporal and spatial coordination of the movements of different articulators. Their point of departure is a specific hypothesis about synchronous programming of lip and tongue movements in the production of Swedish VCV syllables, based on an electromyographic study by McAllister, Lubker, and Carlson (1974). Although the specific hypothesis about synchronous programming was not supported by cinefluorographic data examined in the 1977 paper, the authors nevertheless conclude that the broad concept of interarticulator programming is a viable one that merits further investigation. In support of this conclusion, they cite data from historical phonology and from coordination of phonatory and articulatory activities in speech.

Further evidence in favor of the stronger version of synchronous programming can be found in some recent studies of speech production. Kent and Moll (1975) used cinefluorography to investigate the articulation of consonant

*Also Lund University, Sweden.

Acknowledgment. I am indebted to Katherine Harris, Hajime Hirose, John Chala and Hirohide Yoshioka for comments on an earlier version of this paper. This work was supported in part by NINCDS grant NS-13617 and BRSG grant RR-05596.

clusters beginning with /sp-/ and found that closure for /p/ and release of the constriction for /s/ occurred almost simultaneously, irrespective of linguistic environment. Also using cinefluorography, Gay (1977) noted that, in the first vowel to the stop portion of a V-stop-V sequence, the closing movements of the tongue, jaw, and the primary articulator started almost simultaneously. This finding indicates the possibility of synchronous programming of the movements of different articulators under at least some conditions.

In addition to these examples of temporal coordination, spatial and temporal coordination of different articulators towards achieving a specified goal can be illustrated by the activity of the upper lip, the lower lip and the jaw in the control of vertical lip opening in vowels and occlusion in bilabial stops (Folkins & Abbs, 1975; Hughes & Abbs, 1976). These articulators can be regarded as a coordinated system where the activity of one of them is dependent upon the activity of the others, i.e., if the jaw is constrained so that it cannot move freely to participate in the formation of a labial closure, the upper and lower lips will compensate for the decreased contribution of the jaw to lip closure. Furthermore, similar interrelationships have been observed during vowel production when the jaw or the lips are prevented from moving freely (e.g., Lindblom, Lubker, & Gay, 1979; Riordan, 1977). In these cases, the acoustic characteristics of a vowel remain almost unchanged from the normal condition, indicating that some other articulator must have compensated for the lack of contribution from the jaw or the lips in order to achieve the goal of producing a signal with a specific acoustic structure.

Voiceless obstruent production requires precise temporal control and coordination of several articulatory systems. The tongue, the lips and the jaw are engaged in the formation of the constriction or occlusion; the soft palate is elevated in order to seal off the entrance to the nasal cavity and prevent air from escaping by that route; the vocal folds are abducted in order to prevent glottal vibrations and, by reducing laryngeal resistance to air flow, assist in the buildup of oral air pressure behind the constriction or occlusion. Obstruent production thus provides ample material for investigations of temporal and spatial aspects of interarticulator programming in speech.

The present study was designed to contribute some information on the temporal coordination of laryngeal and oral articulations in the production of Swedish stops. The coordination of these two articulations has proved to be important for the control of aspiration. This study examines how aspiration and its control mechanisms are affected when changes in closure duration and aspiration of a stop are introduced as a result of varying the placement of stress and the number of segments in a word. Although the difference between aspirated and unaspirated voiceless stops is not phonemic in Swedish, when aspiration occurs it serves as one of the cues for the distinction between voiced and voiceless stops, since the former are always unaspirated.

The implications for--and the relation of the present work to--current theories and notions about speech motor control, can be briefly stated as follows.

Much work in speech physiology (see Kent, 1976, for a review) has been carried out within a paradigm where two general questions have dominated: chain versus comb models for motor control of articulation and the role of peripheral feedback in speech production. [The terms "chain" and "comb" are due to Bernstein (1967); in a chain model, the execution of a part of a motor program is triggered by the accomplishment of the preceding part, whereas in a comb model, the various parts are executed independently of each other according to a higher plan.] One limitation in the theoretical approach has been a tendency to subsume the latter question under the former, phrasing the alternatives, as either a chain model incorporating feedback or a comb model without feedback. Of the two remaining alternatives, one is perhaps automatically ruled out, i.e., a chain model without feedback, but the possibility of a comb model incorporating feedback generally has not been exploited, in spite of the wealth of material indicating the existence of peripheral receptors and their general importance in motor control (e.g., Granit, 1970; Matthews, 1972; Sussman, 1972; Wyke, 1967). Another limitation has been an apparent insistence that signals from peripheral receptors must go to higher nervous centers with the resulting problem of apparently inadequate loop time. Another approach would be that information from the periphery goes to lower levels, and there is evidence that such lower levels may play a crucial executive role in integrating signals from higher centers with signals from the periphery. This has been shown for respiratory control (Newsom Davis & Sears, 1970; Sears, 1973), for control of posture and movement (Gottlieb & Agarwal, 1973) and has also been suggested for phonation (Wyke, 1974). Indeed, Denny-Brown (1966) noted that there is no need to postulate a network within the cerebral cortex for detailed cooperation of muscles since it already exists in the spinal segments. Thus some kind of hybrid system might be posited where initiation and goal of a movement are preprogrammed while feedback is used during its execution (e.g., Polit & Bizzi, 1979).

The problem of speech motor control has usually been seen as one of accommodating and coordinating in space and time the articulatory demands for successive segments in the speech chain and studies of coarticulation have generally been directed towards this problem (Daniloff & Hammarberg, 1973; Kent & Minifie, 1977). Since the articulatory units have usually been taken to be more or less identical with the units of linguistic analysis, the temporal resolution necessary in most speech production models has been of the order of magnitude of the segment. A segmental approach has been further encouraged by the fact that the feature representation of segments at a systematic phonetic level, with few exceptions, contains no intrasegmental temporal domain, and such feature representations have often been taken as the input to the speech production apparatus. One of the immediate problems with this approach is to account for the proper sequencing of articulatory movements when these movements do not begin or end at the apparent boundaries between segments (Kent, Carney, & Severeid, 1974). For some classes of speech sounds such as voiceless obstruents, clicks, ejectives, and implosives it is, furthermore, necessary to posit a temporal domain for articulatory movements within one and the same linguistic and/or articulatory unit. This paper discusses one such instance in the production of voiceless stops.

The present experiments were designed to investigate further interarticulator programming in speech--specifically, laryngeal-oral coordination in stop production. Another purpose was to obtain further information on laryngeal

articulatory dynamics in order to evaluate various models and proposals for the control of aspiration in stop consonants. These models will be discussed in more detail below in relation to the results. Some aspects of this work have been discussed previously in L8fqvist (1975, 1976).

METHOD

Laryngeal activity was studied by use of the transillumination technique, also referred to as photoglottography. It is based on the principle that light that enters the subglottic space through the skin from an external light source is modulated when it passes the glottis with variations in glottal opening area, and these modulations can be sensed by a phototransistor placed in the pharynx. Sonesson (1960) improved the technique and applied it to systematic studies of laryngeal activity during phonation. The method has certain limitations, one of which is that the relation between actual glottal opening area and the amplitude of the signal cannot, at present, be calibrated. The amplitude of the glottogram depends, *inter alia*, on the relative position of light source and light sensor and their placement is critical if the signal is to give any useful information. Since conflicting results concerning the accuracy with which the method reproduces actual variations in glottal opening area during phonation have been presented by Coleman and Wendahl (1968) and Harden (1975), it appears unwise to draw any firm conclusions about differences in glottal opening from the glottogram (Hutters, 1976). In spite of these uncertainties, temporal patterns of glottal area changes in obstruent production derived by fiberoptic filming and by simultaneous transillumination of the larynx have proved to be practically identical (L8fqvist & Yoshioka, 1979; Yoshioka, L8fqvist, & Hirose, 1979), indicating that the method appears to provide a realistic picture of the temporal course of the glottal opening. In the present study interest will mainly be focused on temporal aspects of laryngeal articulation.

The light source of the glottograph (LG 900, F-J Electronics) was placed on the skin at the level of the cricothyroid membrane with the light entering the subglottic space from a nearly vertical position. The light sensor was placed in a transparent plastic catheter and introduced into the pharynx through the nose. The subject swallowed the free end of the catheter into the esophagus in order to stabilize the catheter and maintain the transistor in the same position irrespective of articulatory movements. The output from the glottograph was monitored on an oscilloscope and checked for variations in signal quality during the recording session.

To obtain information on oral articulations, simultaneous recordings of oral egressive air flow and intraoral air pressure were made in addition to the glottogram. Air flow was registered via a 2 channel Electro-aerometer (EA 510/2, F-J Electronics) and oral pressure was sampled through a plastic tube inserted into the pharyngeal cavity through the nose and connected to a differential pressure transducer (EMT 33, Siemens-Elema). The glottogram and the aerodynamic signals, along with the signal from a larynx microphone placed at the level of the thyroid cartilage, were recorded on a Mingograph at a paper speed of 100 mm/sec.

MATERIAL AND MEASUREMENTS

The transillumination technique requires a free passage for the light from the glottis to the sensor, thus front vowels and labial and dental consonants are the most suitable linguistic material to use. In the present investigation the following nonsense words were used:

1. 'teten 2. 'tetten 3. 'teteten
4. te'te 5. te'teten 6. tete'teten

All of which represent common Swedish stress patterns. (The apostrophe indicates primary stress.)

The use of a dental stop as representative of all categories of voiceless stops in Swedish was justified by the findings in a pilot study (reported in Löfqvist, 1976) that included the labial and velar places of articulation as well. Although the degree of aspiration in stop consonants varies according to the place of articulation of the stop and the nature of a following vowel (Löfqvist, 1976), no significant differences in laryngeal behavior could be detected among stops with different places of articulation along the parameters of interarticulator programming investigated in the present study and described in more detail below. Variations in the duration of the period of aspiration were therefore assumed to reflect differences in resistance to air flow in the vocal tract after stop release. This, in itself, would explain why the time necessary for the pressure drop across the glottis to reach a level suitable for voicing after stop release would differ according to place of articulation for the stop, and the nature of a following vowel, even if the laryngeal articulation remained the same.

All the test words were placed in the sentence frame "Men se ... igen" and read 20 times from randomized lists by two native male speakers of Swedish.

A general problem in studies of speech physiology is that of defining measurements that are relevant and interesting from the point of view of motor control. Since implosion and explosion in stops are controlled by muscular--and nonmuscular--forces, they were chosen as reference points. Stop closure duration was measured as the interval from the point at which oral pressure started to rise abruptly to the point at which it began to decrease and oral air flow started. Aspiration was taken as the interval between stop release and the onset of glottal vibrations for a following vowel. As indexes of laryngeal articulation, measurements were made of the intervals from stop implosion to the point at which peak glottal opening occurred and from peak glottal opening to release. The point of peak glottal opening is easy to identify, whereas it is almost impossible to determine in the glottogram where the glottis begins to open, cf. Figure 1. In the present study the closing gesture of the glottis generally was found to begin during the closure period and hence no aerodynamic forces can be responsible for its initiation. The point of peak glottal opening must thus be under motor control since it marks the end of the abduction and the beginning of the adduction of the vocal folds; EMG recordings from internal laryngeal muscles have indicated a pattern of reciprocal activation for the posterior cricoarytenoid and the interaryten-

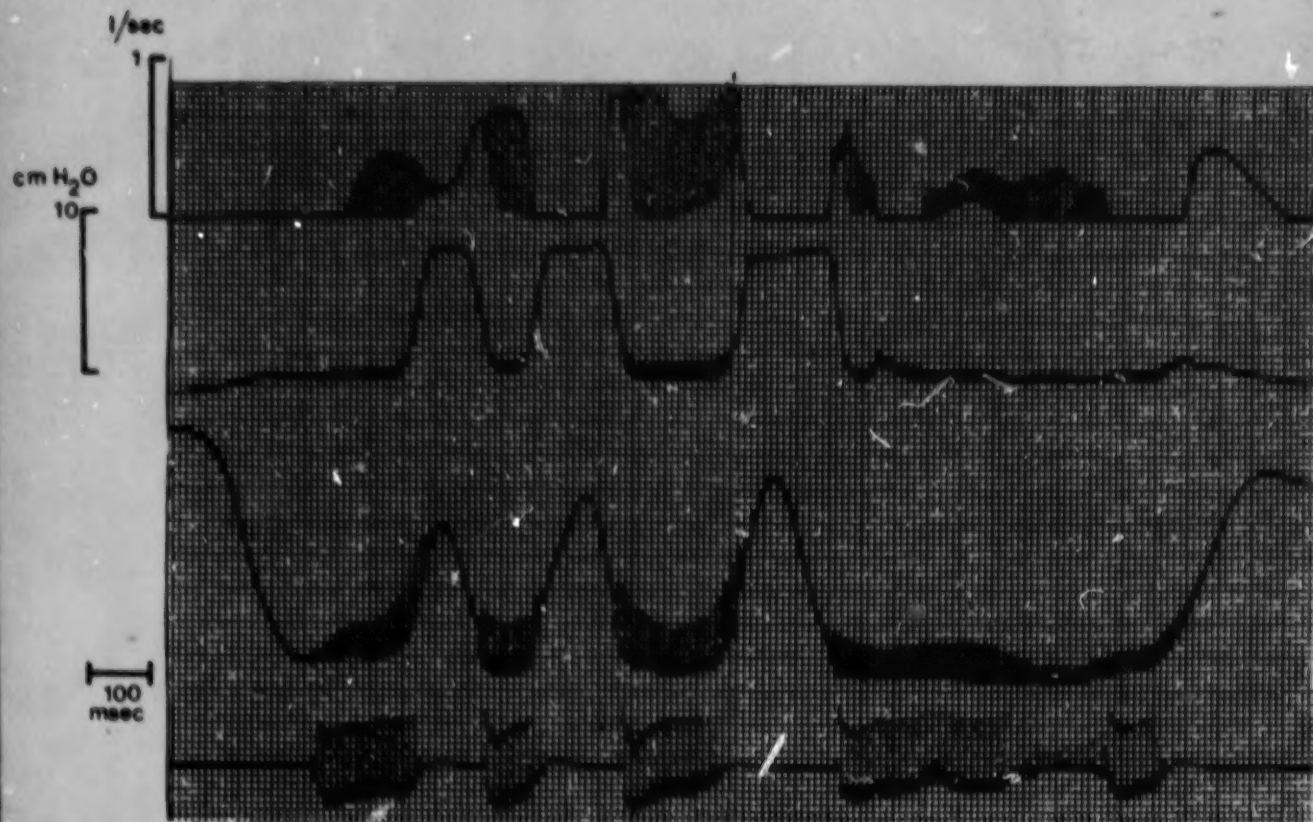


Figure 1. Record of the utterance "Men se 'teten igen." The curves represent from top to bottom oral air flow, oral air pressure, photoglottogram and signal from larynx microphone.

oid muscles in the control of glottal opening in single voiceless obstruents (Hirose, 1976; Hirose, Yoshioka, & Niimi, 1978). This seems to justify the use of peak glottal opening as a reference point in studies of laryngeal articulation in speech.

RESULTS

A sample record of a representative test utterance is presented in Figure 1, and the results of the measurements are summarized in Tables 1 and 2 for the two speakers. The interval from peak glottal opening to release was calculated by subtracting the interval from implosion to peak glottal opening from closure duration and a negative value for this parameter thus indicates that peak glottal opening occurred after stop release. For clarity of exposition and in order to more clearly bring out certain timing relationships some of the parameters are plotted against each other in Figures 2-5.

In two positions, one for each speaker, C_2 in *tete'teten* and C_3 in *'teteten*, respectively, the glottal opening was too small to allow any measurements; hence no measurements of interarticulator timing were made. These positions are not included in the graphs.

Some interspeaker variability is apparent in Tables 1 and 2. Aspiration, closure duration and the interval from peak glottal opening to release are generally longer for speaker 2 than for speaker 1. The difference in closure duration between phonologically long and short stops is greater for speaker 2 and this reflects the different dialects of the speakers. Due to these facts and to others reported below, it was decided not to pool the data but to present the results for each speaker separately.

In Tables 1 and 2, two groups of stops can be identified according to degree of aspiration and closure duration. The first contains stops immediately following a stressed vowel; these have short values for aspiration and will be considered unaspirated. The other group of stops is characterized by longer values for aspiration and contains stops in all other positions except those two where no glottal opening could be found. There is, additionally, a certain relationship between aspiration and closure duration, illustrated in Figure 2. Closure duration is longer for the unaspirated stops and within the group of aspirated stops there is a positive correlation between closure duration and aspiration for speaker 1 but this is less clear for speaker 2.

A similar relation also holds between closure duration and the interval from implosion to peak glottal opening in Figure 3. This interval is generally shorter for unaspirated stops. Among the aspirated ones there is a positive correlation between these two parameters. This indicates, of course, that peak glottal opening tends to occur later during stop closure for aspirated than for unaspirated stops. Within the former group, peak glottal opening occurs later as the duration of stop closure becomes longer.

Closure duration is plotted against the interval from peak glottal opening to stop release in Figure 4. This interval is shorter for aspirated stops and shows no clear correlation with closure duration, whereas for the unaspirated group it increases as closure duration increases.

Table 1

Closure duration, aspiration, and the intervals from implosion to peak glottal opening and from peak glottal opening to release for speaker 1. (msec, n=20)

Word	Segment		Closure duration	Aspiration	Implosion to peak glottal opening	Peak glottal opening to release
'teten	C ₁	x	100	48	100	0
		s	6.9	5.2	5.1	
	C ₂	x	131	11	72	59
		s	6.7	3.4	10.6	
'tetten	C ₁	x	97	115	101	-4
		s	7.5	5.3	8.9	
	C ₂	x	158	2	85	73
		s	10.7	2.4	14.8	
'teteten	C ₁	x	96	113	94	2
		s	6.5	4.7	9.4	
	C ₂	x	106	12	64	42
		s	8.7	4.6	9.0	
	C ₃	x	76	26	67	9
		s	9.2	6.3	8.2	
te'te	C ₁	x	79	24	67	12
		s	10.3	5.3	10.3	
	C ₂	x	99	51	89	10
		s	7.7	7.7	7.8	
te'teten	C ₁	x	73	22	63	10
		s	8.8	3.4	7.9	
	C ₂	x	100	42	77	23
		s	8.9	7.5	8.3	
	C ₃	x	121	11	59	62
		s	12.3	4.6	11.4	
tete'teten	C ₁	x	78	27	71	7
		s	8.8	4.7	7.2	
	C ₂	x	57	16		
		s	7.7	4.2		
	C ₃	x	92	39	84	8
		s	7.2	5.9	8.1	
	C ₄	x	119	10	61	58
		s	9.3	4.3	8.8	

Table 2

Closure duration, aspiration, and the intervals from implosion to peak glottal opening and from peak glottal opening to release for speaker 2. (msec, n=20)

<u>word</u>	<u>Segment</u>		<u>Closure duration</u>	<u>Aspiration</u>	<u>Implosion to peak glottal opening</u>	<u>Peak glottal opening to release</u>
'teten	C ₁	x	116	39	89	27
		s	9.3	5.8	9.0	
	C ₂	x	159	23	72	87
		s	9.9	5.3	7.5	
'tetten	C ₁	x	112	35	84	28
		s	6.4	4.3	8.5	
	C ₂	x	234	22	36	148
		s	14.3	3.4	5.5	
'teteten	C ₁	x	117	42	97	20
		s	8.5	6.1	8.8	
	C ₂	x	135	20	65	70
		s	13.7	5.0	9.0	
	C ₃	x	79	21		
		s	5.8	3.9		
te'te	C ₁	x	113	38	88	25
		s	9.3	7.1	9.0	
	C ₂	x	105	57	95	10
		s	8.1	6.8	4.9	
te'teten	C ₁	x	113	39	89	24
		s	9.7	7.5	8.9	
	C ₂	x	101	46	86	15
		s	7.5	7.1	6.8	
	C ₃	x	148	23	65	83
		s	7.2	4.7	5.5	
tete'teten	C ₁	x	99	36	78	21
		s	10.8	5.7	7.7	
	C ₂	x	90	30	61	29
		s	10.1	3.9	7.6	
	C ₃	x	98	48	83	15
		s	7.7	7.0	5.9	
	C ₄	x	143	20	65	78
		s	5.9	4.4	4.4	

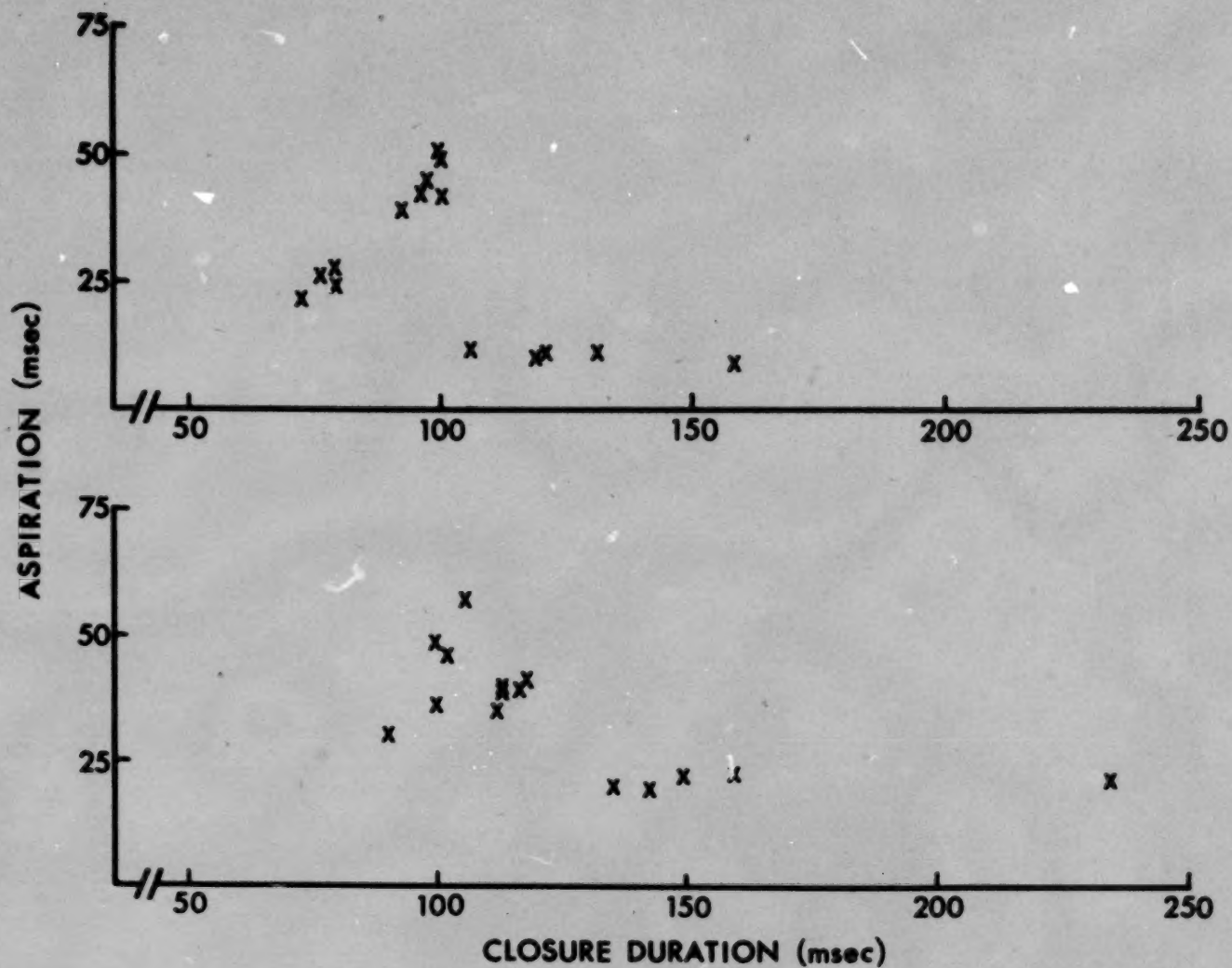


Figure 2. Aspiration plotted versus closure duration for speaker 1 (top) and speaker 2 (bottom).

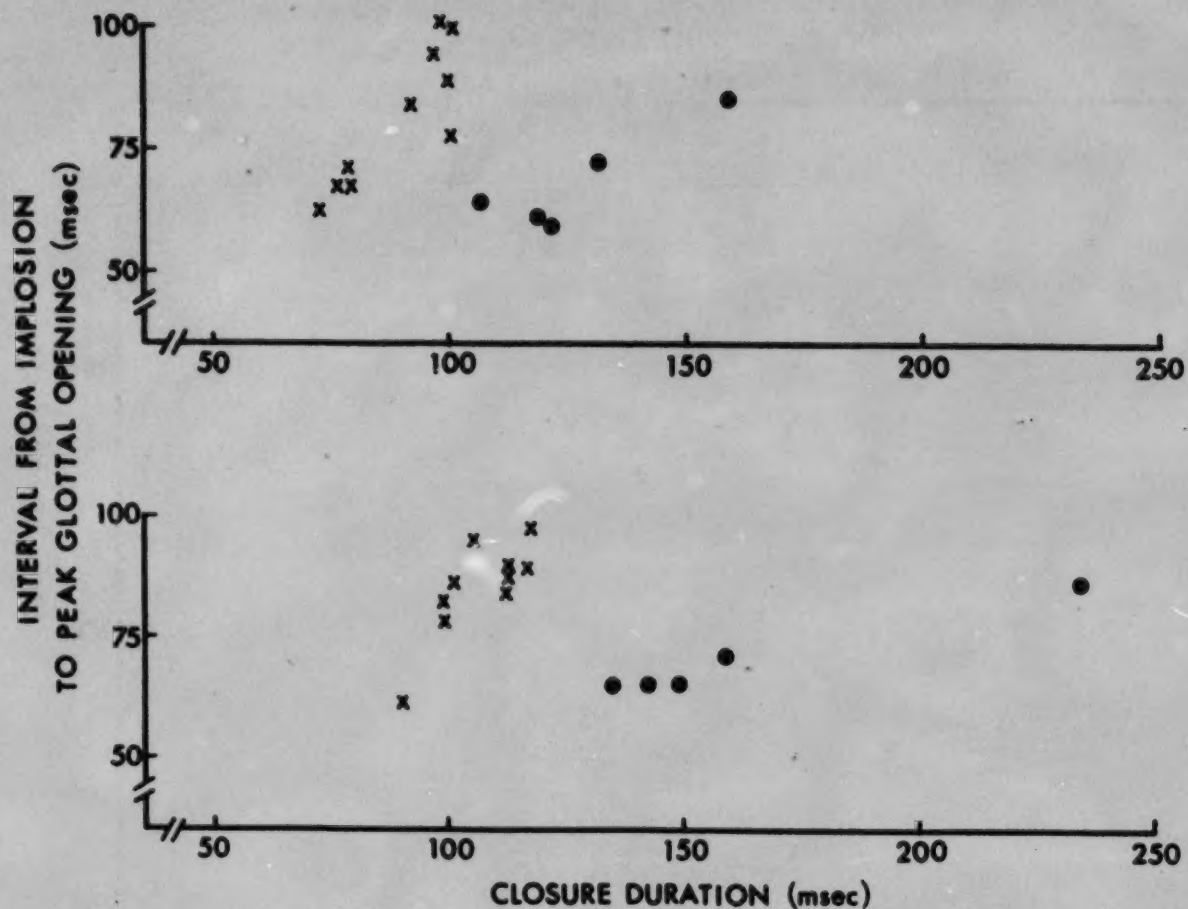


Figure 3. The interval from implosion to peak glottal opening plotted versus closure duration for speaker 1 (top) and speaker 2 (bottom); aspirated stops are denoted by X and unaspirated by filled circles.

Figure 4. The interval from peak glottal opening to release plotted versus closure duration for speaker 1 (X) and speaker 2 (filled circles).

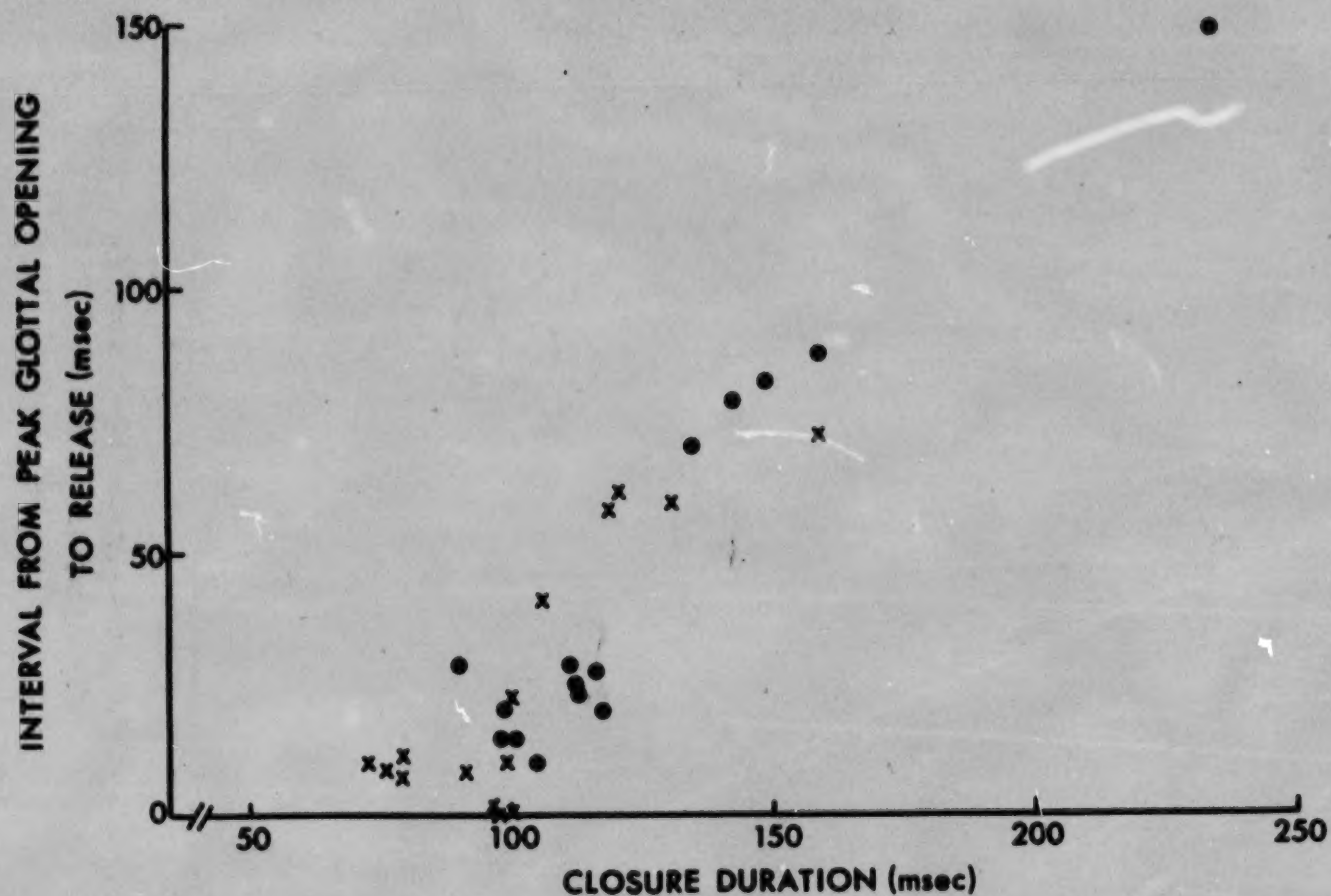
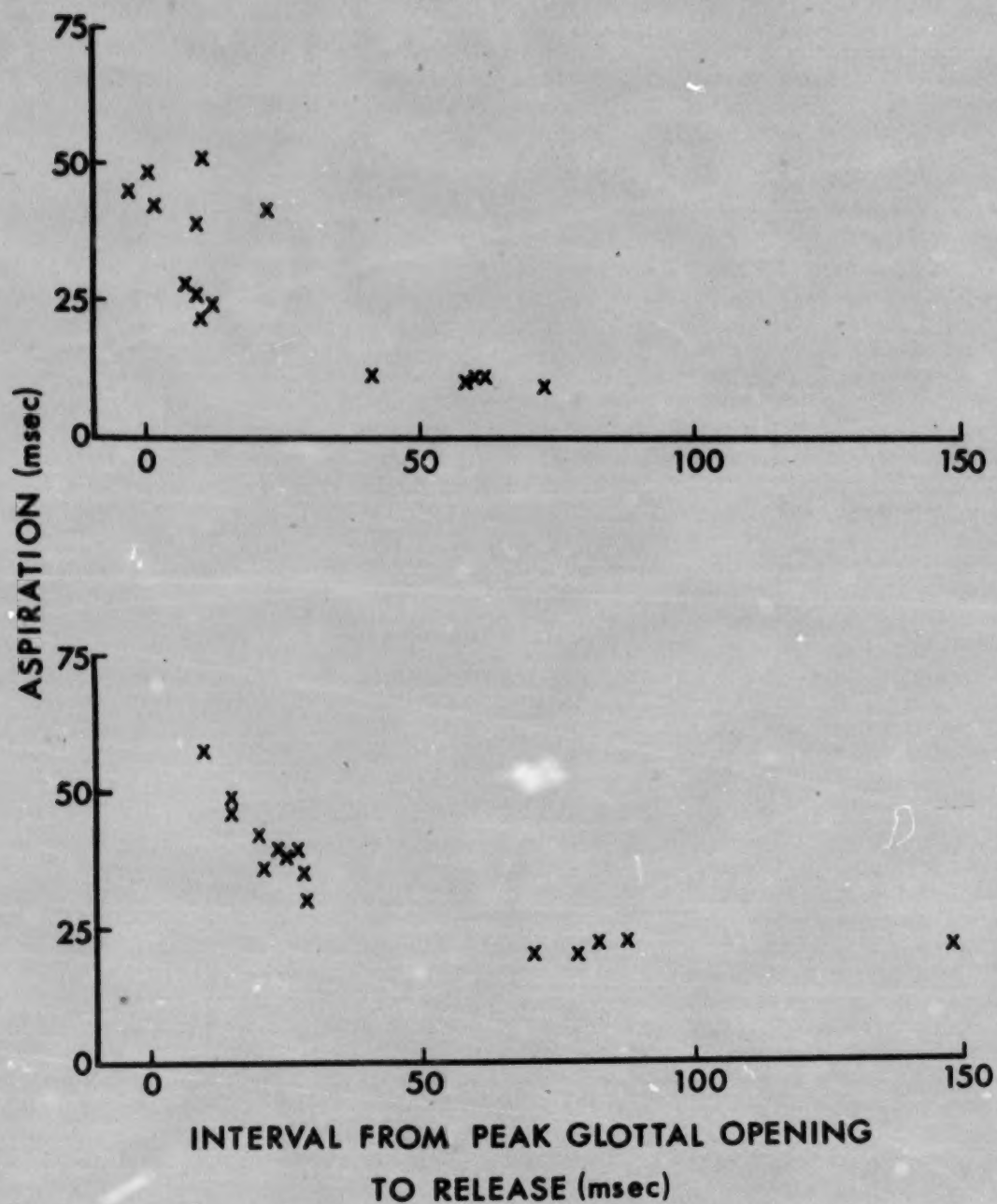


Figure 5. Aspiration plotted versus the interval from peak glottal opening to release for speaker 1 (top) and speaker 2 (bottom).



We turn to the relation between aspiration and the interval from peak glottal opening to release presented in Figure 5. As noted above, this interval is shorter for aspirated stops. Within this group there is a negative relationship between the two parameters for speaker 2 but not for speaker 1. A closer inspection of the results for this speaker reveals that the degree of aspiration is here related to stress and that degree of stress separates the data for aspirated stops into two subgroups. In one of them, stops in syllables with primary stress, aspiration is between 40 and 50 msec irrespective of the duration of the interval from peak glottal opening to stop release. The other subgroup contains stops in syllables with secondary stress where aspiration is about 25 msec and the interval between peak glottal opening and release about 10 msec.

This speaker variability is further illustrated in comparisons of peak glottal opening in different stops. Due to the technical limitations discussed above, no quantitative measurements of glottal opening have been made; but in order to get some notions about this parameter, the conservative approach of only comparing stops within the same utterance was adopted, and within the utterance the stops were ranked according to peak glottal opening area, defined as the height of the glottogram above baseline. One might argue that external factors influencing the glottogram are more likely to remain stable within the same utterance. The ranking procedure also avoids, to some extent, the problem of possible non-linearities in the relation between actual glottal opening area and the amplitude of the glottogram. These rankings were compared with other parameters and the comparisons showed different trends for the two speakers. For speaker 1 the size of peak glottal opening covaried with stress degree. For speaker 2, on the other hand, glottal opening was always closely related to the duration of the oral closure and increased with it. If we try to interpret these relationships in terms of interarticulator timing, glottal opening was larger the longer the interval from implosion to peak glottal opening for speaker 1, whereas for speaker 2 it was larger the shorter the same interval, with some exceptions. These findings indicate that the relation between peak glottal opening and aspiration is not necessarily direct in Swedish but the technical limitations should be kept in mind as well as the fact that no comparisons were made across utterances and that aspiration is non-phonemic in Swedish.

DISCUSSION

The results of the present study emphasize once more the temporal precision achieved in the coordination of oral and laryngeal articulations in obstruent production. They also indicate the possibility and existence of interspeaker variability in the production of signals with similar acoustic structure. Furthermore, they point to the need for studying stop production within a broader linguistic framework that takes into account the distinctive function of aspiration, closure duration and voicing in signaling phonological contrasts. The reason why no clear glottal opening could be detected in two positions in spite of the fact that the stops in these positions were clearly voiceless, is presumably that both these positions, although not the same for the two speakers, are very weakly stressed. Hence it may not be necessary to maintain all aspects of the distinction between voiced and voiceless stops. These positions are exceptions in that short closure durations are found with short periods of aspiration and the short closure perhaps did not allow any

appreciable laryngeal articulatory gesture.

Compared with other studies of laryngeal articulation during stop production in languages as diverse as Danish (Frøkjær-Jensen, Ludvigsen, & Rischel, 1971), English (Sawashima, 1970), French (Benguerel, Hirose, Sawashima, & Ushijima, 1978), Hindi (Dixit, 1975; Kagaya & Hirose, 1975), Icelandic (Pétursson, 1976), Japanese (Sawashima & Niimi, 1974), Korean (Kagaya, 1974; Kim, 1970), Mandarin (Iwata & Hirose, 1976) as well as with that on Swedish by Lindqvist (1972) the results from the present investigation show both agreement and disagreement.

A general feature of laryngeal articulation for voiceless obstruents that emerges from all these investigations is that the vocal folds seem to be constantly moving in what can be described as a single ballistic opening and closing gesture. In only one case is there any evidence, for Hindi voiceless unaspirated stops reported by Kagaya and Hirose (1975), of the glottis opening and maintaining a static position until the closing gesture starts, and in this single case it is not a regular feature but only occurs in a limited number of tokens. Thus, laryngeal articulation appears to be a continuous gesture and this seems to be the case also for clusters of voiceless obstruents (Löfqvist, 1977; Pétursson, 1977) where, under some conditions, the continuous movement takes the form of two successive opening and closing gestures (Löfqvist, 1978; Pétursson, 1978). The same ballistic opening and closing pattern can also be observed in utterance initial and utterance final position (Lindqvist, 1972; Löfqvist, 1976, 1977) and it seems worthwhile to incorporate this characteristic feature of laryngeal articulation into a general model of laryngeal function in speech.

The results for Swedish stops presented above indicate that the timing of this laryngeal gesture in relation to supraglottal events is the decisive factor in the control of aspiration and, as suggested by, among others, Lisker and Abramson (1964, 1971), and Rothenberg (1968), different temporal coordinations of these two articulations would seem to explain and account for most existing features of pre-aspiration and post-aspiration in stop consonants. If the glottal opening gesture starts prior to oral closure, pre-aspiration results, as in Icelandic. If it starts at implosion and peak glottal opening occurs early during stop closure, the stop is unaspirated, whereas if peak glottal opening occurs late during closure, post-aspiration results. When the glottal opening gesture starts at the release, the result is a voiced aspirated stop as in Hindi. All these patterns can thus be viewed as arising from different timing relationships between the laryngeal opening and closing gesture and the oral closing and opening gesture in stop production.

Even if this temporal relation appears to be of primary importance, it is conceivable that other parameters of the glottal gesture, such as velocity of glottal movement and size of glottal opening, might also be independently controlled and used in obstruent production. No direct information is available on velocity and the data on size are somewhat uncertain since neither the transillumination technique nor motion pictures of the larynx taken via a fiberoptic can be accurately calibrated. For the latter technique this is due to the fact that the larynx may move up and down during speech (cf. Ewan & Krones, 1974) and thus the distance from the glottis to the lens will vary.

According to the studies referred to above, variations in peak glottal opening tend to occur mainly as a function of whether the stop is aspirated or not, and peak glottal opening is generally larger in the former case. In view of the rather limited number of subjects investigated thus far, it is not yet possible to determine whether these variations are speaker specific or language specific. The former seems to be the case for the Swedish data reported here, and does not seem unreasonable if one considers the way in which a child may learn to produce voiceless obstruents, since different strategies are available, cf. below. Even if differences in peak glottal opening were a regular phenomenon in the production of different stop categories, it should be noted that, in the published studies, these size differences always appear to be accompanied by the timing differences discussed above. Thus, it appears unwarranted to claim that the size difference is more basic than the timing difference.

It seems logical to view changes in timing and size of glottal opening as two interacting strategies in stop production. Their combined use in the production of a voiceless unaspirated stop can thus manifest itself as an early timing of peak glottal opening during stop closure along with a comparatively small glottal opening. In this case, both will contribute to an adducted glottis at stop release. More generally, variations in both of these dimensions can thus be regarded as different ways of achieving a certain degree of glottal opening at release which, as was noted by Kim (1970), is one of the chief determinants of degree of aspiration in voiceless stops, at least in those instances where peak glottal opening precedes the release. For speaker 2 there is, in Figure 5, a neat inverse relation between aspiration and the interval from peak glottal opening to release, but this is not so clear for speaker 1 where the size of peak glottal opening, related to stress, would seem to play a certain role.

If this claim by Kim (1970) thus appears to be true, this is not necessarily the case for another claim made in the same paper: That size of (peak) glottal opening, and not the time at which the glottis begins to close, is directly controlled in stop production. The time at which the glottis begins to close is clearly not invariant, since we saw above in Figure 3 that the location of peak glottal opening occurs at different times in relation to both implosion and release for aspirated and unaspirated stops; within the aspirated group, peak glottal opening is consistently delayed in relation to implosion as a function of closure duration. We should also note that peak glottal opening and glottal opening at release are, in general, not identical. Although a more rigorous experimental test of these two theories may not be readily designed, a consideration of the broader framework of laryngeal articulatory dynamics would seem to make the timing theory a more reasonable one. Timing thus appears to be the basic way in which the articulatory system solves the problem of controlling glottal opening at release, and thereby, the onset of glottal vibrations in relation to the explosion.

We can further illustrate the use of different strategies in producing voiceless aspirated and unaspirated stops and how they are used in different languages if we also take closure duration into account as an independent parameter. In Figure 3 it is apparent that when aspirated and unaspirated stops in Swedish have about the same closure duration, the difference between the two groups in the interval from implosion to peak glottal opening is

generally larger than when they differ widely in closure duration. The same phenomenon can be seen in Icelandic (Löfqvist & Pétursson, 1978) where aspirated and unaspirated voiceless stops have about the same closure duration and thus show a large difference in the interval from implosion to peak glottal opening. This obviously reflects the tighter requirement of timing peak glottal opening early during the closure if closure duration is short and could, at least for the Icelandic data, be seen in less variance in the interval from implosion to peak glottal opening for unaspirated stops. If closure duration is long, as is Swedish unaspirated stops, there is more time for the glottis to return to a position suitable for voicing and less precision is required in interarticulator timing. It should also be noted that in at least some languages other than Swedish (Danish, English, Hindi, Korean) closure duration is generally longer for unaspirated than for aspirated voiceless stops.

Several interacting strategies can thus be used in the production of voiceless unaspirated stops--among them an early timing of peak glottal opening and an increase in closure duration. Within the timing framework adopted here, it is possible to give a hypothetical but phonetically plausible account of the emergence of pre-aspiration in stop consonants and why it never seems to co-occur with post-aspiration. In order to avoid post-aspiration, an early timing of peak glottal opening during the closure can be used. In this process, the coordination of glottal opening and oral implosion may be more or less synchronous; and if glottal opening precedes oral closure, an audible noise will occur that might eventually develop into a regular phonologic pattern. In fact, pre-aspiration has been reported as a regular feature of some Swedish dialects and then always for voiceless stops without post-aspiration. A unified account of these phenomena would seem possible only within a timing theory.

The same framework can also provide an account for some observations on children's productions of obstruents and how they evolve with age. Studies of voice onset time in children's productions of American English stops (Kewley-Port & Preston 1974; Zlatin & Koenigsknecht, 1976; Gilbert, 1977) show that children under 2 years of age mainly produce stops with short voicing lag and do not make any consistent difference between voiced and voiceless utterance initial stops on the basis of VOT. Later, the VOT values begin to show the bimodal distribution characteristic of adult speakers. Zlatin and Koenigsknecht (1976) present some suggestive results on the range of VOT for initial stops in the speech of children 2 and 6 years of age and adults. The 2-year-olds and the adults show opposite patterns for range with the 6-year-olds falling in between. The adults have a large range of VOT for voiced stops but a much smaller range for voiceless stops, whereas the 2-year-olds have a larger range for voiceless than for voiced stops. Presumably, the range reflects several things, such as the ability to coordinate and control laryngeal and oral articulations, the extent to which phonological patterns have been learned and internalized, and the precision required by the phonological system. The age-related range variation can presumably be ascribed to different factors for children and adults.

The children's consistent production of short voicing lag stops can most likely be accounted for along the lines given by Kewley-Port and Preston (1974), i.e., a closed glottis during closure or a closing of the glottis

before release would result in short VOT values due to aerodynamic factors. The larger variation among voiceless stops would reflect difficulties with the necessary temporal coordination required for their production. The smaller variation found for voiceless stops in adult speech would be due to their mastering the articulatory timing involved and perhaps also to a phonologic feature of American English that voiceless initial stops be produced with values of voice onset time within a restricted range. The greater variability for voiced stops would reflect the fact that prevoicing is not phonemic in American English and hence a greater variability is allowed by the linguistic code. Similar age-related data on Swedish stops are not available at present.

The same pattern also emerges from a study of patients suffering from apraxia of speech (Freeman, Sands, & Harris, 1978). The greater timing requirements in obstruent production show up in the inability of apraxic speakers to produce consistent patterns of VOT and their successive return to a more normal distribution after a period of therapy and recovery.

A more general question and one that will have to await a more definite answer is how this interarticulator coordination is achieved. Clearly, the initiation of the glottal closing gesture cannot be triggered by afferent impulses signalling the drop in oral pressure at release, as has been suggested. Since the start of the glottal closing gesture occurs at different times in relation to both implosion and release, it appears impossible to design a simple peripheral trigger mechanism that would account for the coordination of the oral and laryngeal articulatory gestures in this case, specifically in view of the fact that the motor events producing the movements occur about 50 - 100 msec prior to the movements themselves. The most reasonable conclusion seems to be that they are preprogrammed as a whole in order to produce the acoustic variations that, according to Swedish phonology, occur when stress and the number of segments in the test word are changed. One might add that chain, and comb, models in general have been directed towards the sequencing of successive units but not the sequencing of articulatory movements within a unit. Moreover, it remains unclear what the relevant peripheral events triggering successive units might actually be.

The articulatory movements of the glottis during obstruent production appear to be rather stereotypic and mostly consist of an opening and closing gesture. The same gesture also often occurs in utterance initial and utterance final position (Lindqvist, 1972; Löfqvist, 1976, 1977; Sawashima, Hirose, Ushijima, & Niimi, 1975), although it appears to be more common in utterance final position. In utterance initial position, the glottis may first close from a respiratory position and then execute the articulatory gesture, whereas in utterance final position the closing part of the gesture is executed before the glottis returns to a respiratory position. The laryngeal gesture would thus seem to be an inherent feature in the production of voiceless stops and fricatives and perhaps also clusters of voiceless obstruents (Löfqvist, 1978).

Some experimental paradigms that can help in further clarifying the nature of laryngeal-oral coordination in obstruent production are currently being explored. One involves studying this coordination across different speaking rates. Another is to apply sudden loads to the jaw and the lips and observe whether a perturbation of the oral articulators results in a concomi-

tant change in the glottal movements when the load is applied at implosion or release and at different times during these phases. This would presumably indicate any dependency of laryngeal articulation on oral articulatory movements whereas dependencies in the other direction cannot be as readily elucidated. A useful theoretical framework for these studies is that of coordinative structures developed primarily by Russian scholars (Bernstein, 1967; Gelfand, Gurfinkel, Fomin, & Tsetlin, 1971; Turvey, 1977). Designed to cope with the number of degrees of freedom to be directly controlled, this theory views motor coordination in terms of constraints among muscles or groups of muscles that have been set up for the execution of specified movements. The experiments briefly outlined above might indicate whether such a concept of coordinative structures is a valid one for laryngeal-oral coordination in obstruent production.

REFERENCES

- Benguereel, A.-P., Hirose, H., Sawashima, M., & Ushijima, T. Laryngeal control in French stop production: A fiberoptic, acoustic and electromyographic study. Folia Phoniatrica, 1978, 30, 175-198.
- Bernstein, N. The co-ordination and regulation of movements. New York: Pergamon, 1967.
- Coleman, R., & Wendahl, R. On the validity of laryngeal photosensor monitoring. Journal of the Acoustical Society of America, 1968, 44, 1733-1735.
- Daniloff, R., & Hammarberg, R. On defining coarticulation. Journal of Phonetics, 1973, 1, 239-248.
- Denny-Brown, D. The cerebral control of movement. Liverpool: Liverpool University Press, 1966.
- Dixit, R. P. Neuromuscular aspects of laryngeal control: With special reference to Hindi. Unpublished doctoral dissertation, The University of Texas at Austin, 1975.
- Ewan, W. G., & Krones, R. Measuring larynx movement using the thyrohyoidmeter. Journal of Phonetics, 1974, 2, 327-335.
- Folkins, J. W., & Abbs, J. H. Lip and jaw motor control during speech: Responses to resistive loading of the jaw. Journal of Speech and Hearing Research, 1975, 18, 207-220.
- Freeman, F. J., Sands, E. S., & Harris, K. S. Temporal coordination of phonation and articulation in a case of verbal apraxia: A voice onset time study. Brain and Language, 1978, 6, 106-111.
- Frøkjær-Jensen, B., Ludvigsen, C., & Rischel, J. A glottographic study of some Danish consonants. In L. Hammerich, R. Jakobson, & E. Zwirner (Eds.), Form and substance. Copenhagen: Akademisk Forlag, 1971.
- Gay, T. Articulatory movements in VCV sequences. Journal of the Acoustical Society of America, 1977, 62, 183-193.
- Gelfand, I. M., Gurfinkel, V. S., Fomin, S. V., & Tsetlin, M. L. (Eds.), Models of the structural functional organization of certain biological systems. Cambridge, Mass.: MIT Press, 1971.
- Gilbert, J. H. A voice onset time analysis of apical stop production in 3-year olds. Journal of Child Language, 1977, 4, 103-110.
- Gottlieb, G. L., & Agarwal, G. C. Coordination of posture and movement. In J. E. Desmedt (Ed.), New developments in electromyography and clinical neurophysiology (Vol. 3). Basel: Karger, 1973.
- Granit, R. The basis of motor control. London: Academic Press, 1970.
- Harden, R. J. Comparison of glottal area changes as measured from ultrahigh-

- speed photographs and photoelectric glottographs. Journal of Speech and Hearing Research, 1975, 18, 728-738.
- Hirose, H. Posterior cricoarytenoid as a speech muscle. Annals of Otology, Rhinology and Laryngology, 1976, 85, 334-342.
- Hirose, H., Yoshioka, H., & Niimi, S. A cross language study of laryngeal adjustment in consonant production. Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, University of Tokyo, 1978, 12, 61-71.
- Hughes, O. M., & Abbs, J. H. Labial-mandibular coordination in the production of speech: Implications for the operation of motor equivalence. Phonetica, 1976, 33, 189-201.
- Hutters, B. Problems in the use of the photoelectric glottograph. Annual Report of the Institute of Phonetics, University of Copenhagen, 1976, 10, 275-312.
- Iwata, R., & Hirose, H. Fiberoptic acoustic studies of Mandarin stops and affricates. Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, University of Tokyo, 1976, 10, 47-60.
- Kagaya, R. A fiberoptic and acoustic study of the Korean stops, affricates and fricatives. Journal of Phonetics, 1974, 2, 161-180.
- Kagaya, R., & Hirose, H. Fiberoptic, electromyographic and acoustic analysis of Hindi stop consonants. Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, University of Tokyo, 1975, 9, 27-46.
- Kent, R. D. Models of speech production. In N. Lass (Ed.), Contemporary issues in experimental phonetics. New York: Academic Press, 1976.
- Kent, R. D., Carney, P. J., & Severeid, L. R. Velar movement and timing: Evaluation of a model for binary control. Journal of Speech and Hearing Research, 1974, 17, 470-488.
- Kent, R. D., & Minifie, F. D. Coarticulation in recent speech production models. Journal of Phonetics, 1977, 5, 115-134.
- Kent, R. D., & Moll, K. L. Articulatory timing in selected consonants. Brain and Language, 1975, 2, 304-323.
- Kewley-Port, D., & Preston, M. S. Early apical stop production: A voice onset time analysis. Journal of Phonetics, 1974, 2, 195-210.
- Kim, C-W. A theory of aspiration. Phonetica, 1970, 21, 107-116.
- Lindblom, B., Lubker, J., & Gay, T. Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation. Journal of Phonetics, 1979, 7, 147-161.
- Lindqvist, J. Laryngeal articulation studied on Swedish subjects. Quarterly Progress and Status Report of the Speech Transmission Laboratory, Royal Institute of Technology, Stockholm, 1972, 2/3, 10-27.
- Lisker, L., & Abramson, A. A cross language study of voicing in initial stops: Acoustical measurements. Word, 1964, 20, 384-422.
- Lisker, L., & Abramson, A. Distinctive features and laryngeal control. Language, 1971, 47, 767-785.
- Löfqvist, A. On the control of aspiration in Swedish. Paper presented at 8th International Congress of Phonetic Sciences, Leeds, 1975. Working Papers of the Phonetics Laboratory, Department of General Linguistics, Lund University, 1975, 12, 99-106.
- Löfqvist, A. Closure duration and aspiration for Swedish stops. Working Papers of the Phonetics Laboratory, Department of General Linguistics, Lund University, 1976, 13, 1-39.
- Löfqvist, A. Artikulatorisk programmering - några data kring svenska obstruentkombinationer. Nysvenska Studier, 1977, 57, 95-104.

- Löfqvist, A. Laryngeal articulation and junctures in the production of Swedish consonant sequences. In E. Gårding, R. Bruce, & R. Bannert (Eds.), Nordic prosody. (Travaux de l'Institut de Linguistique de Lund XIII). Lund: Department of Linguistics, 1978.
- Löfqvist, A., & Pétursson, M. Swedish and Icelandic stops - a glottographic investigation. In J. Weinstock (Ed.), The nordic languages and modern linguistics 3. Austin: The University of Texas at Austin, 1978.
- Löfqvist, A., & Yoshioka, H. Laryngeal articulation in Swedish obstruent clusters. In J. Wolf & D. Klatt (Eds.), Speech communication papers presented at the 97th meeting of the Acoustical Society of America. New York: The Acoustical Society of America, 1979.
- Lubker, J., McAllister, R., & Lindblom, B. On the notion of inter-articulator programming. Journal of Phonetics, 1977, 5, 213-226.
- Matthews, P. B. C. Mammalian muscle receptors and their central actions. London: Edward Arnold, 1972.
- McAllister, R., Lubker, J., & Carlson, J. An EMG study of some characteristics of the Swedish rounded vowels. Journal of Phonetics, 1974, 2, 267-278.
- Newsom Davis, J., & Sears, T. A. The proprioceptive reflex control of the intercostal muscles during their voluntary activation. Journal of Physiology, 1970, 209, 711-738.
- Pétursson, M. Aspiration et activité glottale. Phonetica, 1976, 33, 169-198.
- Pétursson, M. Timing of glottal events in the production of aspiration after /s/. Journal of Phonetics, 1977, 5, 205-212.
- Pétursson, M. Jointure au niveau glottal. Phonetica, 1978, 35, 65-85.
- Polit, A., & Bizzi, E. Characteristics of motor programs underlying arm movements in monkeys. Journal of Neurophysiology, 1979, 42, 183-194.
- Riordan, C. Control of vocal tract length in speech. Journal of the Acoustical Society of America, 1977, 62, 998-1002.
- Rothenberg, M. The breath-stream dynamics of simple-released-plosive production. (Bibliotheca Phonetica, Vol. 6). Basel: Karger, 1968.
- Sawashima, M. Glottal adjustments for English obstruents. Haskins Laboratories Status Report on Speech Research, 1970, SR-21/22, 187-200.
- Sawashima, M., Hirose, H., Ushijima, T., & Niimi, S. Laryngeal control in Japanese consonants, with special reference to those in utterance initial position. Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, University of Tokyo, 1975, 9, 21-26.
- Sawashima, M., & Niimi, S. Laryngeal conditions in articulations of Japanese voiceless consonants. Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, University of Tokyo, 1974, 8, 13-17.
- Sears, T. A. Servo control of the intercostal muscles. In J. E. Desmedt (Ed.), New developments in electromyography and clinical neurophysiology (Vol. 3). Basel: Karger, 1973.
- Sonesson, B. On the anatomy and vibratory pattern of the human vocal folds. Acta Oto-laryngologica (Supplementum 156), 1960.
- Sussman, H. What the tongue tells the brain. Psychological Bulletin, 1972, 77, 262-272.
- Turvey, M. Preliminaries to a theory of action with reference to vision. In R. Shaw & J. Bransford (Eds.), Perceiving, acting and knowing. Hillsdale, N.J.: Erlbaum, 1977.
- Wyke, B. Recent advances in the neurology of phonation: Phonatory reflex mechanisms in the larynx. British Journal of Disorders of Communication, 1967, 2, 2-14.

- Wyke, B. Laryngeal myotatic reflexes and phonation. Folia Phoniatrica, 1974, 26, 249-264.
- Yoshida, H., Lofqvist, A., & Hirose, H. Laryngeal adjustments during American English consonant cluster production. In J. Wolf & D. Klatt (Eds.), Speech communication papers presented at the 57th meeting of the Acoustical Society of America. New York: The Acoustical Society of America, 1979.
- Zlatin, M. A., & Koenigsknecht, R. A. Development of the voicing contrast: A comparison of voice onset time in stop perception and production. Journal of Speech and Hearing Research, 1976, 19, 93-111.

THE BEGINNINGS OF SPEECH

Michael Studdert-Kennedy*

INTRODUCTION

Man's life is diverse. The range of habitats, natural and man-made, to which he has adapted is incomparably wider than that of any other species. This is so because there evolved in man capacities for rapid cultural evolution to augment the lengthy biological processes of adaptive radiation. These capacities have permitted him to create new and unpredictable patterns of behavior in the face of both old and new contingencies. The nature of these capacities is quite unknown. But we can be sure that language is among them, and that an understanding of its biology would take us a long way toward understanding the history of man and of the earth during the past 10,000 years.

Unfortunately, "...the development of human speech represents a quantum jump in evolution comparable to the assembly of the eucaryotic cell" (Wilson, 1975, p. 556). Whatever the lost links in phyletic evolution since the first hominids diverged from the apes, presently living species offer few analogies and even fewer homologies with language. In fact, the most fruitful approaches to its biology seem to be those that have been followed for many years by developmental psycholinguists (for reviews, see Brown, 1973; Dale, 1976; Ferguson & Slobin, 1973) and by students of neurophysiology (e.g., Lenneberg, 1967; Lenneberg & Lenneberg, 1975; Whitaker & Whitaker, 1976): first, study of its ontogeny, with particular attention to similarities within and across language communities; second, study of its pathology in childhood and adult disorders.

The present chapter makes no attempt to review the vast, resulting literature. Instead it undertakes to examine, critically, several tempting analogies with language in the great apes and in the song-learning of certain birds. Analogies often have the heuristic value of leading us to look at familiar facts from a fresh viewpoint. Moreover, they may be instructive even if they prove to be false.

*Also Queens College and Graduate Center, City University of New York.

Acknowledgment: Most of the work on this chapter was done while the author was on sabbatical leave as a visiting scientist at the Center for Interdisciplinary Research, University of Bielefeld, Germany, and a participant in a nine-month conference of anthropologists, biologists, and psychologists on Early Development in Animals and Man (September 1, 1977 through May 31, 1978). The chapter will be published in G. B. Barlow, K. Immelmann, M. Main, and L. Petrinovich (Eds.), Behavioral Development: The Bielefeld Interdisciplinary Project. New York: Cambridge University Press, 1980.

[HASKINS LABORATORIES: Status Report on Speech Research SR-58 (1979)]

THE NATURE OF LANGUAGE

If we compare language with other animal communication systems, we are struck by its breadth of function. The flashing white rump of the fallow deer denotes alarm; the "peep" of the squirrel monkey indicates that it is alone and wishes it wasn't; the "song" of the chaffinch informs the interested listener of its species, sex, local origin, personal identity and readiness to breed or fight. Even the elaborate "dance" of the honey bee merely conveys information about the direction, distance and quality of a nectar trove. But language can convey information about all these matters and many more besides. In fact, it is the peculiar property of language to set no limit on the possible topics of reference.

More exactly, no language consists of a finite number of sentences. This may be demonstrated by formal proof (Chomsky, 1956), or by the persuasive calculation that a single rendering of all grammatical English sentences of up to, say, twenty words in length would last longer than the history of the earth (Miller, Galanter, & Pribram, 1960, p. 146). In fact, no normal speaker of a language--no matter how limited his vocabulary or tedious his conversation--speaks by rote or constructs an utterance by drawing its components from a store of ready-made phrases.

How does language achieve this openness or productivity? There are several crucial features to its design (Hockett, 1960). First, language is learned: It develops under the control of an open, rather than of a closed genetic program (Mayr, 1974). Transmission of the code from one generation to the next is therefore discontinuous: Each individual recreates the system for himself. There is ample room here for creative error--probably a central factor in the evolution of language and in the constant process of change that all languages undergo (Kiparsky, 1968). One incidental consequence of this freedom is that the universal properties of language (whatever they may be) are largely masked by the surface variety of the several thousand languages now spoken in the world, not to mention their thousands of dialects and idiolects.

A second condition of productivity is that linguistic signals are arbitrary. With a few onomatopoeic exceptions, only by coincidence does a sign share any property with its referent. Of course, many other animal signals are arbitrary: the courtship rituals of the great-crested grebe, the red spot of the courting stickleback, the flush of a shamed human. But under the surface of such instances, some unknown physiological necessity is at work. These are not the arbitrary signs of convention by which bird, oiseau, Vogel and uccello are equivalent. Notice that if signs were iconic rather than arbitrary, the number of possible referents would be limited by the signaling organism's physical capacity to represent or depict.

A third, closely related condition of productivity is that signals are discrete rather than analog. To be precise, signals are perceived as discrete, even if they are not physically separable. Here again, if signals were not categorized by the receiver and if changes of meaning required changes of degree along some continuous scale, the number of possible signals

would be limited by the number of possibly and perceptibly variable dimensions of the signal.

A final condition of productivity, and the one to which we will give most attention, is that language has two hierarchically related levels of structure: Its signal elements are combined according to two more-or-less independent systems of rules. At the lower level of each language, the phonology or sound system, a small set (usually between 20 and 60) of meaningless phonemes (consonants and vowels) is specified, together with rules for their combination into morphemes (meaningful units which, for present purposes, we may treat as roughly equivalent to words). These are the rules that permit a vast, if not infinite, lexicon to be constructed by permutation and combination of a few dozen "alphabetic" units.

At a second level of structure, that of syntax, are specified the rules for combining words into meaningful sentences. These are the rules that permit us to predicate relations among objects or events. Central to the syntax of every known language are "recursive rules" by which a sentence may be treated as a component in another sentence. This capacity to embed a sentence within a sentence means that the set of all possible sentences in a language is infinite (Chomsky, 1956). Moreover, it is through this device that we can extend our communicative reach by constructing complex, sentential "names" for referents not represented in our lexicon, a trick already in the armory of many 3-year-olds: "I want the one Mary's got" (Limber, 1973). Incidentally, it is this central, inventive (though commonplace) use of language that Premack (1976, p. 15) thinks it "absurd" to expect of the chimpanzee.

IMPLICATIONS OF DUAL STRUCTURE

We begin to apprehend the importance of a dual structure, if we imagine a language with only one level, say that of sound (cf. Liberman & Studdert-Kennedy, 1978). Such a language would consist of meaningless elements (perhaps consonants and vowels) combined into lexical items, a set of "words" each with a different referent. Its users would presumably be confined to ostensive definition. For even if they were able to conceive of absent objects ("The bear we met yesterday") or abstract ideas ("The solar year") and were able to construct, from their phonetic resources, new lexical items to refer to them, they would be quite unable, lacking discursive speech, to establish the new meanings with their fellows. It is only by means of syntax that we are able to deploy old (known) words into new (previously unknown) statements--such as those that define new words. In short, rules for syntactic structure are a sine qua non of linguistic productivity.

The lack of a sound structure, on the other hand, would be less crippling. For, even if we were to replace every word in the lexicon with an arbitrary number (as might be done if the lexicon were stored in a computer), the syntactic structure of any particular utterance would be preserved despite the total loss of phonetic equivalences. (It is for this reason that linguists sometimes describe a language as an abstract system of communication, independent of its medium of expression.) Each lexical item would then be a totally distinct sign, lacking any systematic physical relation to any other. Of course, the number of such irreducible, holistically distinct

signals that humans are capable of recalling, producing and identifying at even a moderate rate--let alone the 50 bits/second typical of much speech--is certainly small, and it is not surprising that most vertebrate communication systems dispose of no more than 10 to 40 signals (Wilson, 1975, p. 183). However, a small lexicon does not preclude a productive syntax. That is why Premack (1976) and Rumbaugh (1977) saw no need for a formational structure in the visual symbols they devised for their pongid pupils.

Nonetheless, having granted that phonological (or word formational) structure is not, in principle, necessary for productive language, we must next acknowledge that every known language does, in fact, display it. The "extra" level of sound structure--which perhaps was prior to syntax in phyletic evolution, as it is in ontogeny--must therefore fulfill some function.

That function, as we have already suggested, is to facilitate the formation of a lexicon. Whether or not the lexical, or "naming," function is at the root of language, as is sometimes argued (e.g., Lancaster, 1968), most linguistic communities do have--in addition to their everyday lexicon of several thousand words--large, more-or-less specialized vocabularies, crucial to their cultural elaboration of the environment. This is as true of "primitive" peoples, such as the Hanunoo of the Philippines with their vast inventories of flora and fauna (Levi-Strauss, 1968) as of a modern industrial society with its proliferation of technical terms and subculture jargon. Thus, the seemingly trivial discovery that an essentially unlimited lexicon could be constructed from a small "alphabet" of sounds may have been the catalyst that set linguistic development in motion by providing an interface between man's intellect and his peripheral anatomic structure (Liberman, 1970; Mattingly, 1975). Certainly, it is at the level of the signaling system (that is, of speech) rather than of the abstract syntactic and semantic structure, that we find the clearest traces of biological adaptation, and it is therefore primarily with speech that the following sections are concerned.

THE SIGNALING SYSTEM

The sounds of any language can be viewed as the product of a sound source and a resonant filter. The sound source is usually either the "voice" produced by rapid pulsing of the vocal cords (as in the final sounds of "be" and "do"), the hiss of air blown through a narrow constriction (as in the initial and final sounds of "safe" and "thrush") or both (as in the final sounds of "leave" and "bees"). The resonant filter is the vocal tract, that is, the cavities of the pharynx, mouth, and nose.

The pulsing of the vocal cords at fundamental frequencies of roughly 90 to 250 Hz for males, 150 to 350 Hz for females and somewhat higher for small children, yields a signal rich in harmonic frequencies (multiples of the fundamental). Relatively slow variations in fundamental frequency over the course of an utterance yield the characteristic melody or intonation of speech. Taken with systematic variations in intensity, rate and rhythm, this melody is the basis of speech prosody, and plays an important role in communicating the emotional tone of an utterance, as well as, to some extent, its syntactic structure (e.g., question, statement, imperative). To the

unfamiliar listener (whether infant or foreigner) the slow variations of prosody are probably more salient than the rapid patter of consonant-vowel syllables. But it is primarily by syllables that the distinctively linguistic (lexical and syntactic) information is carried. That, incidentally, is why writing systems encode phonetic segments, but not prosody.

For the most part, this distinctively linguistic information is conveyed by systematic variations in the "tuning" of the vocal tract. The curved column of air in the tract, like that in an Alpine horn, resonates in characteristic frequency bands (or formants) when set in motion by air from a vibrating source, with the result that some of the source frequency components are amplified, while others are attenuated. If we vary the size and shape of the resonating tract by shifting the relative positions of the articulators, especially the tongue, lips, jaw, and soft palate, the resulting shifts in the formants yield the various sound spectra characteristic of particular phonetic segments. The reader may find it instructive to monitor the position and shape of his tongue as he runs it around the vowel triangle: eat, it, et, at, ah, ought, oot.

THE SOUND PATTERN OF LANGUAGE

Here we must introduce the concept of a sound system or phonology. Each language forms its words from a relatively small "alphabet" of distinctive phonetic segments, termed phonemes. These are its consonants and vowels, and in English there are about 35 of them, depending on dialect. The phonemes are not chosen randomly. Each may be described in terms of the small set of binary features (usually, a dozen or so) deployed in a particular language. The phonemes may then be classified according to their shared features and the resulting classes contrasted with one another on the basis of their feature oppositions. A basic division, observed in every language, is between consonants, formed by a more or less complete constriction of the vocal tract, and vowels, formed with a relatively open tract. From their contrastive combination is formed the fundamental unit of all spoken language, the consonant-vowel syllable. It is the repeated opening and closing of the tract and the consequent repetitive frequency and amplitude modulation, or syllabic beat, that establishes the characteristic rhythms of human speech.

We may draw further contrasts among the phonemes (Figure 1). For example, in English we may draw contrasts between voiced (/b,d,v,z/) and voiceless (/p,t,f,s/), between continuant (/s,f,z,v/) and stop (/t,p,d,b/), between constriction at the alveolar ridge behind the upper front teeth (/s,z,t,d/) and constriction at the lips (/f,v,p,b/). Taken together these eight phonemes, formed from three binary contrasts, constitute a little system within the larger system of English phonology.

The particular selection of features used in any language is largely determined by phonetic drift over time and by a complex of historical and social forces. But the universal stock of phonetic features is presumably constrained by human anatomy and physiology: It must be drawn from the (as yet unspecified) intersection of what we can articulate with what we can perceive. The goal of much work (e.g., Jakobson, Fant, & Halle, 1963; Chomsky & Halle, 1968; Ladefoged, 1971) has been to define the smallest set of universal features (perhaps fewer than 20) that will include all features that

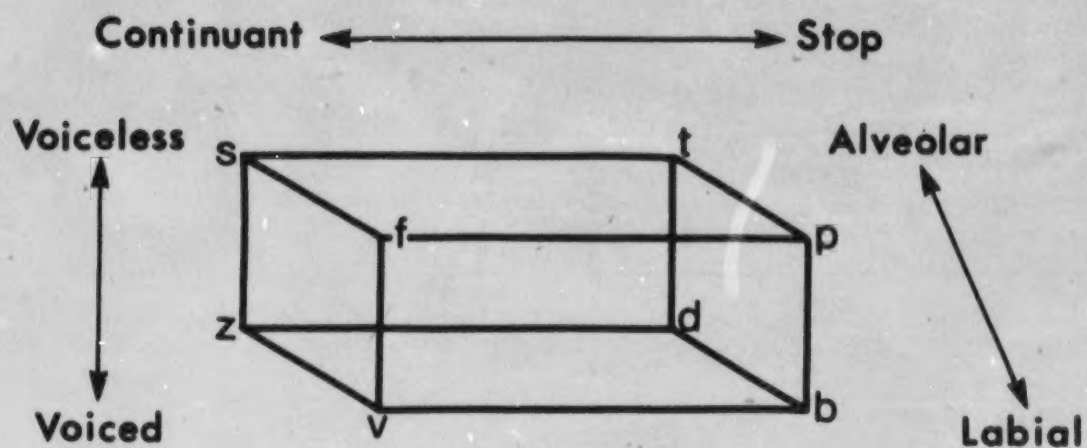


Figure 1. A three-dimensional binary feature space, excerpted from the multi-dimensional feature space that describes the English phonological system.

may be distinctive in any language.

But there is more to the phonology of a language than the structure of its phonemic system. Each language also disposes of more-or-less elaborate rules for combining phonemes into words: These are the rules of its syllable structure. For example, in English the basic syllable structure can be represented as: (C)(C)(C)V(C)(C)(C)(C)(C), where C = consonant, V = vowel and parentheses indicate that the slot may or may not be filled. Thus, the simplest syllable is an isolated vowel. But in most syllables the required vowel is preceded by up to three consonants and followed by up to five consonants (the latter only in a few rare words such as "triumph'st") (Abercrombie, 1967).

Moreover, there are strict limits on the permissible consonant clusters. For example, in English, if two obstruents (stops or fricatives) occur together, the voicing of the second must match the voicing of the first. Accordingly, English words may begin with sp-, st-, or sk-, but not with sb-, sd- or sg-. Hence, too, the plurals in -s or -z, (apes, lions), the present indicatives in -s or -z (she raps, she loves) and the past in -t or -d (rapped, loved). A subsidiary rule states that, if the two obstruents are formed by closure at roughly the same point in the vocal tract, a neutral vowel (the so-called schwa) must be inserted between them, giving the plural, roses, the present indicative, she kisses, the past, she hated. Most normal children, growing up among English speakers, have unconsciously learned these rules by the age of about six, and therefore have no difficulty in forming the correct plurals, presents and pasts of words they have never heard before (Berko, 1958).

The point of this example is to make clear that very much more is required to learn the sound structure of a language than the capacity to listen and to imitate. In fact, as we shall see below, even within its first year of life, the infant has begun to discover and apply rules.

THE FUNCTION OF PHONETIC FEATURES

We have defined features up to this point in articulatory terms. In part, this is because precise acoustic description, drawing on spectrographic analysis, has proved intractable. But it is principally because articulation is, in fact, prior to the acoustic signal. Indeed, it has been plausibly argued that the feature structure of spoken language was primarily a solution to the problem of getting high speed articulatory performance out of low speed articulatory machinery (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). The feature structure permits a shift from one phoneme to the next by a change of no more than one or a few articulatory features. The value of articulatory ease is attested by the universal phenomenon of assimilation. Every language has many rules by which certain sounds or classes of sounds take on features of neighboring sounds, permitting a "lazier," and so more rapid, articulation. For example, the final n of the prefix syn- (synthesis, synechdoche) becomes m in sympiosis and sympathy, taking on the labial articulation of the following consonant. Similarly, normally voiced l, sounded with laryngeal pulsing in light, takes on the voiceless feature of s in a word such as slight.

Of course, a gain for the speaker may be a loss for the listener. It is precisely such shifts in articulation and the consequent subtle shingling of the acoustic properties of neighboring phonemes that have thwarted attempts at automatic speech recognition and given rise to the central problems for a theory of speech perception. Parallel (or co-) articulation of consonant and vowel in the integral ballistic gesture of the syllable (Stetson, 1952) gives rise to an acoustic signal in which the cues to a particular phoneme vary widely as a function of context and in which the boundaries between successive phonemes are obliterated. The tempting model that language might have been expected to offer for the division of motor behavior into "natural" units is thus a mirage. The units are not to be found either in the articulation or in the acoustic signal. The problem of segmentation appears to be solved by perceptual fiat. Not surprisingly, this has encouraged theorists of speech perception to invoke exotic perceptual mechanisms such as analysis-by-synthesis (Stevens & Halle, 1967; cf. Liberman et al., 1967) and "dedicated" property or feature detecting devices (see below).

Perhaps specialized perceptual mechanisms have indeed evolved to match the specialized motor mechanisms. There is strong evidence in vocal tract morphology, in tongue and lip innervation, in mechanisms for breath control during speech, and so on, that extensive adaptations for speaking did occur (Lenneberg, 1967; Lieberman, 1972; Du Brul, 1977). Perhaps these and matching perceptual adaptations (including specialized sensorimotor processes for imitation) underlie the evolution of language. However, once the capacity for language had evolved, man was able to deploy it in another mode. What is interesting is that, when he does so, as in American Sign Language, the formal structure of the system remains largely unchanged.

AN ALTERNATIVE SIGNALING SYSTEM: MANUAL SIGN LANGUAGE

Visual and tactile finger-spelling, like alphabetic and syllabic writing, are parasitic on speech: They simply transpose its units into another modality. However, some visual languages are independent of spoken language: for example, the sign languages of the American Plains Indians (West, 1960), of the Australian aborigines (Umiker-Sebeok & Sebeok, 1977), and of countless deaf communities in the various countries of the world (Stokoe, 1974). The signs of these languages do not necessarily correspond to the words of any particular spoken language, nor do the rules for their combination follow the syntax of any spoken language.

Consider, as an example, since it has been the most extensively studied, American Sign Language (ASL or Ameslan). Ameslan is a derivative of the French sign language introduced by Gallaudet to the U.S. in 1817: Users of Ameslan today are said to understand French SL better than British SL—evidence for the independence of sign and spoken languages. The first dictionary of Ameslan (Stokoe, Casterline, & Croneberg, 1965) contains over 2,000 signs. Many of them seem iconic, but usually not until one knows what they mean—just as one may not recognize the metaphor in, say, "The road runs west" until it is pointed out. Other signs are indexical: Pronouns, for example, are often formed by pointing. However, pointing and pure pantomime are rare. The overwhelming majority of signs are arbitrary or, if once iconic, have now lost much of their iconicity (Frishberg, 1975).

Signs may use one or two hands, and may vary along at least three orthogonal dimensions: configuration, position within the signing area (a rough circle around head and chest, centered below the chin), and movement. Stokoe et al. (1965) have analyzed the values along these dimensions into some 55 "cheremes," a number well within the phonemic count of many spoken languages. Later work (e.g., Battison, 1974; Klima & Bellugi, 1979; Lane, Boyes-Braem, & Bellugi, 1976) has demonstrated that formational rules govern the possible combinations of cheremes into signs, just as the phonological rules of a language govern the combination of phonemes into words. Finally, Ameslan has now been shown to possess a richly inflected grammar and a syntax, that is, a set of rules governing the spatial and temporal ordering of signs into sentences (Klima & Bellugi, 1979; see also Siple, 1979). In short, Ameslan displays all the distinctive properties of a human language including a dual pattern of form and syntax.

The significance of this recent work on Ameslan is twofold. First, it underlines the link between hand and mouth, and the likely importance of a rapid, informationally-dense signaling system for efficient linguistic communication, a point to which we return below (see also Studdert-Kennedy, 1977). Second, it demonstrates the abstract nature of the capacities underlying language development. So far as we know, no other animal has developed a capacity for essentially equivalent communication in two different sensorimotor systems.

THE GREAT APES

Recent successes in training apes to communicate by means of artificial symbol systems (Premack, 1976; Rumbaugh, 1977) or a natural sign language (Ameslan) (Gardner & Gardner, 1969, 1975; Terrace, Pettito, & Bever, 1976a, 1976b; Patterson, 1978) have shown that the cognitive, representational and perhaps even linguistic capacities of chimpanzees and gorillas, though vastly inferior, are nonetheless very much closer to those of man than was once thought. Given the tight genetic relation between man and chimpanzee (King & Wilson, 1975) and their very different ecologies, one may wonder whether these apparently similar behavioral capacities in man and ape may not be homologous capacities derived by genetic transmission from a common ancestor.

Unfortunately, the degree of similarity and its evolutionary implications are difficult to assess because none of the supposedly linguistic behaviors of the apes seems to occur naturally. All have required intervention by animals of another species in the form of systematic operant conditioning. This is particularly striking in the work of Premack (1976) and Rumbaugh (1977) where chains of behavior are established by direct shaping and primary reinforcement of hundreds of responses with food, drink, bodily contact and so on. For the signing chimpanzees, such as Washoe (Gardner & Gardner, 1975) and Nim (Terrace et al., 1976a, 1976b), the social reward of trainer approval is more usual. Nonetheless, even here the fundamental training procedure has been operant shaping and molding of specific behaviors. In other words, language learning in the great apes does not proceed without the establishment of stimulus-response contingencies.

By contrast, the human infant is apparently disposed to learn language even in the absence of specific response shaping and reinforcement. While it

too may require the generalized social reinforcement of a partner's attention, the infant does not require shaping and reinforcement of particular responses. On the contrary, as Brown (1973) has remarked, parents tend to reinforce the truth value, but not the form of their children's utterances. In other words, language develops in spite of the absence of narrowly-defined stimulus-response contingencies.

Particularly striking in this context is the recent work of Feldman, Goldin-Meadow, and Gleitman (1977) on the spontaneous development of signing in deaf children. They studied six deaf children, over an age span of 1,5 through 4,6 1/2, whose parents were following the "oralist" practice recommended by some authorities in the U.S.A. These authorities believe that signing to congenitally deaf children lowers their motivation to lip-read and articulate English; they therefore urge parents and siblings of such children to avoid all gestures, formal or informal. According to Feldman et al., the families of their six subjects were largely successful in following this practice.

The procedure of the study was to videotape each child playing and passing time with its mother and the experimenter during several standardized home visits. In the course of playing with the toys and games introduced by the experimenter every child devised its own "home-signs," that is, a characteristic set of motor-iconic gestures to refer to objects, actions, predicates. Moreover, each child gradually began to combine these signs into two-, three-, and even six-sign sequences, creating its own semantically-based syntax, including systematic deletion rules of the kind observed in a normal hearing child's "telegraphic" speech. This last point is particularly interesting, since telegraphic signing was not produced by the adults conversing with the children any more than is telegraphic speech under normal circumstances. The authors end their lengthy analysis with the conclusion that "...there are significant internal dispositions in humans that guide the language acquisition process" (Feldman et al., 1977).

There is, of course, no evidence for such dispositions in the ape. This argues that the cognitive capacities now being discovered in the apes are general rather than specifically linguistic. The adaptive functions of these capacities are not always obvious. For example, how does the wild chimpanzee use its capacity to symbolize? Or is this capacity perhaps a "neo-phenotype" (Kuo, 1976; Miller, 1980), an item of general behavioral plasticity, not normally deployed, but available for use in the face of the right selective pressures?

Another general capacity, impressively displayed in the recent language projects, does have obvious utility, namely, the capacity to learn a new motor response by observation and imitation. This requires that the animal, first, be able to parse perceived behavior into action components, and second, have sensorimotor connections by which the parsed patterns may be mapped into motor commands (cf. Terrace et al., 1976a, p. 21). Field observations attest to the role of imitation in the young chimpanzee's learning to fish for termites, for example, or to build its nest (van Lawick-Goodall, 1971).

Yet a third chimpanzee capacity, essential to linguistic communication, has recently been demonstrated by Premack and Woodruff (1978)--the attributing

of "intention" to the behavior of another organism. Here again, the capacity, whatever its linguistic worth, obviously contributes to the development of social intelligence. In fact, laboratory studies of ape "language acquisition" probably have more to teach us about the evolutionary origins of mind than of language. Certainly, as Limber (1977) suggests, conversational chimpanzees may offer an experimental approach to the study of relations between language and thought (for example, does naming facilitate problem-solving?), but the focus would then be on thought rather than on language. For insight into the origins of language, the frank analogues of birdsong may have more to offer than the possible homologues of ape signs.

THE SONG BIRDS

Unlike observational learning of other motor behavior, vocal learning can have no value beyond its use in communication. The analogous appearance of vocal learning in both man and bird is therefore of special interest (Marler, 1970, 1975; Nottebohm, 1970, 1975). Indeed, Marler has proposed as "...a significant evolutionary step toward...the strategy of speech development of *Homo sapiens*..." the emergence of "...new sensory mechanisms for processing speech sounds..." as well as "...neural circuitry...to modify patterns of motor outflow so that sounds generated can be matched to preestablished auditory templates" (Marler, 1975, pp. 32-33). As we shall see, the evidence for "new sensory mechanisms" or "auditory templates" in humans is weak, but there is good evidence for specialized sensorimotor processes.

Templates and Feature Detectors

Birds (and other animals). Species-specific templates were proposed by Marler (1963, p. 233) and Konishi (1965) to account for the fact that many songbirds prefer to learn the songs of their own species. Even if they are deprived of conspecific song during the sensitive phase, and are exposed to the songs of closely related species, they tend not to learn them (e.g., Marler & Peters, 1977).

The form of these templates, "...lying in the auditory pathway" (Marler, 1975, p. 26) has never been specified. However, presumably they would consist of networks of specialized neurons tuned to particular properties of the species' song. Cortical neurons sensitive to changing frequencies were reported for the cat ("miaow cells") by Whitfield and Evans (1965). Cells tuned to species calls have been reported for the bullfrog (Frishkopf & Goldstein, 1963; Capranica, 1965), the squirrel monkey (Wollberg & Newman, 1972), several species of echo-locating bat (Neuweiler, 1977) and the starling (Leppelsack & Vogt, 1976).

Humans. A possible analogy between species-specific call or song detectors and phonetically relevant, acoustic feature detectors was not lost on students of speech perception (e.g., Abbs & Sussman, 1971; Liberman et al., 1967; Studdert-Kennedy, 1974). The feature detector promised to solve at a single blow a variety of problems in speech perception, including that of syllable segmentation. Moreover, the notion of feature with its roots in ethology, linguistics and pattern recognition was attractive to biologically-inclined students of language, looking for signs of an innate acquisition device (e.g., Stevens, 1975). Unfortunately, the several lines of evidence

and speculation seem to have converged on an error.

The story begins with the phenomenon of categorical perception (Eimas, 1963; Liberman et al., 1967; Studdert-Kennedy et al., 1970). Early work with speech synthesizers showed that it was a simple matter to construct acoustic tokens of opponent phonetic types by varying a single acoustic parameter. For example, by varying the interval between plosive release and the onset of laryngeal pulsing, that is, voice onset time (VOT), one could construct a continuum of, say, a dozen tokens ranging in equal acoustic steps from /ba/ to /pa/, or from /da/ to /ta/.

If listeners were asked to identify these tokens, they showed a strong tendency to call any particular stimulus by the same name (e.g., /ba/) every time they heard it. There were few, if any, ambiguous tokens. Furthermore, if they were asked to discriminate between neighboring pairs of tokens, they tended to do badly if they judged the two tokens to be members of the same phoneme class, but well, if they judged the tokens to be members of opponent phoneme classes—even though the acoustic interval between pairs was identical in the two cases. This phenomenon, dubbed "categorical perception," seemed to be a useful process for speech perception. After all, one cannot afford to judge a word to be more-or-less "bat" or more-or-less "pat." One must categorize it instantly as one or the other: Classification is a crucial process in phonetic perception.

The next event in the story was the demonstration by Eimas, Siqueland, Jusczyk, and Vigorito (1971), using a non-nutritive sucking habituation procedure, that one-month- and four-month-old infants could discriminate between two tokens differing by 20 msec along a voice onset time continuum, providing they were tokens that adults normally classified as different phonemes. But the infants could not discriminate between tokens that adults normally classified as the same phoneme. Similar results for a variety of synthetic speech continua were reported in due course for infants growing up in other language communities (see Eimas, 1975, for a review).

The suspicion that these results reflected categorical perception, mediated by specially tuned, innate feature detectors, was not easy to resist—particularly since the phylogenetic emergence of such detectors might then be the evolutionary step that carried hominids from a graded to a categorical communication system (cf. Marler, 1975). The hunt for independent evidence of such detectors operating in human adults began, and, by 1973, Eimas and Corbit were able to report apparent success. They modified a procedure with a long history in visual studies: adaptation. The paradigm is simple enough. For example, prolonged fixation of a red patch of light adapts or fatigues a red detector cell and relatively sensitizes its opponent green detector cell, so that upon looking at a white screen, the viewer sees a relatively unsaturated green patch the same shape as the red adaptor. Related effects in form and tilt also occur. Such effects have frequently been taken as evidence for the operation of opponent feature detectors.

Eimas and Corbit (1973) asked listeners to categorize members of a synthetic voice onset time continuum and demonstrated that the perceptual boundary between voiced and voiceless categories along that continuum was shifted by repeated exposure to (that is, adaptation with) either of the

endpoint stimuli: There was a decrease in the frequency with which stimuli close to the original boundary were assigned to the adapted category and a consequent shift of the boundary toward the adapted stimulus. They took the effect to be evidence for the operation of an opponent feature detecting system. Several dozen studies over the next five years replicated the effect on several other synthetic speech continua. (For reviews, see Ades, 1976; Cooper, 1975; Eimas & Miller, 1978.)

Thus, the chain of inference and speculation from percept to detector was complete. Unfortunately, each link in the chain has proved weak. First, several studies have shown that categorical perception is not peculiar to speech, or even to audition. For example, Pastore et al. (1977) demonstrated categorical perception of critical flicker, with a sharp boundary at the flicker-fusion threshold. Second, other studies (e.g., Carney, Widin, & Viemeister, 1977) have demonstrated that the degree of categorical perception varies with the experimental method used to measure it: Listeners can be trained to hear a supposedly categorical continuum noncategorically or to shift category boundaries from one point on a VOT continuum to another. Finally, cross-language studies have found that speakers of different languages may place phonetic boundaries at different points along the same acoustic continuum, demonstrating that acoustic-phonetic categories are determined linguistically by language experience rather than neurophysiologically by innate feature-detecting devices. (For a review of cross-language studies, see Strange & Jenkins, 1978.)

The demise of categorical perception as a specialized phonetic process also cuts the other links in the chain. Thus, instances of what appears to be infant categorical perception will doubtless find a straightforward account in terms of auditory psychophysics, similar to that developed for the adult case. In fact, Pisoni (1977) has already developed such an account for voice onset time.

By the same token, we no longer need opponent process feature detectors to account for a general psychophysical phenomenon--particularly since there are quite other grounds for doubting the opponent detector model. Most obvious is the model's lack of behavioral or neurological motivation. For, while the facts of additive color mixture and retinal neurophysiology make an opponent detector account of after-effects entirely plausible, the facts of perceived stop consonant onset and cochlear neurophysiology certainly do not. However, an adequate discussion of speech adaptation is well beyond the scope of this chapter, and it must suffice to remark that plausible accounts of the effects in terms of stimulus range (Rosen, in press), auditory contrast (Simon & Studdert-Kennedy, 1978) or other more general processes (Remez, 1979) have already begun to appear.

We must conclude that we now have no evidence for the operation in speech perception of specialized sensory mechanisms analogous to the auditory templates postulated for certain songbirds.

Lateralization and the Sensorimotor Device

Birds. One of the most remarkable discoveries in recent years is the lateralization of neural function in birdsong (Nottebohm, 1971, 1972, 1977)--

at present, the only securely attested instance of lateralized behavior outside man (although see Dawson, 1977, and Petersen, Beecher, Zoloth, Moody, & Stebbins, 1978). The typical songbird syrinx, as instantiated by that of the canary (Nottebohm, 1977), has two independently innervated and functionally separate halves. Sections of the right and left halves (or of their innervating hypoglossal branches) have very different effects: Right side sections lead to the loss of no more than 0 to 15% of pre-operative song syllables, while left side sections lead to a 90 to 100% loss. Similar effects of peripheral lesion have been observed in the chaffinch (Nottebohm, 1971, 1972) and the white-crowned sparrow (Nottebohm & Nottebohm, 1976). For the canary, Nottebohm (1977) has also traced motor pathways from the syrinx to the associated brain structures: Unilateral brain lesions indicate that the left hemisphere contributes radically more to song control than does the right.

All these effects are motor, and no perceptual lateralization has been demonstrated. However, it is of interest that the principal motor control center lies next to the telencephalic auditory projection, where processes involved in establishing the species-specific song template are believed to occur. Indeed, it was Nottebohm's (1970) original notion that lateralization might be associated with complex learned behavior. This view has been thrown into question by the discovery of peripheral lateral equipotentiality in the orange-winged Amazon parrot (Nottebohm, 1976), a bird well-known for its vocal plasticity, and of left lateralization in the domestic fowl (Youngren, Peek, & Phillips, 1974), a bird of equally well-known vocal stereotypy. Nonetheless, current research on the canary is charting neural links between the two centers, in an attempt to isolate the sensorimotor connection, presumably essential to song learning (Kelley & Nottebohm, 1979).

Humans. It has been known for many years that the left cerebral hemisphere contributes more to language function than the right, in most normal humans. The bulk of our knowledge comes from studies of aphasia, induced by stroke, tumor or gunshot wound (e.g., Jenkins, Jimenez-Pabon, Shaw, & Sefer, 1975; Penfield & Roberts, 1959) and, more recently, from studies of "split-brain" patients, whose cerebral hemispheres have been surgically separated by section of the connecting pathways for relief of epilepsy (e.g., Zaidel, 1978a, 1978b). The latter condition permits an investigator to assess the linguistic capacities of each hemisphere independently.

Of particular interest, in light of the bird song findings, is that left hemisphere specialization seems to be primarily for control of the articulatory apparatus and for perceptual analysis of spoken words into their phonetic segments. The human larynx and its associated articulatory structures (tongue, velum, jaw) are bilaterally innervated, but unilaterally controlled. Thus, "verbal apraxia," or aphasic disturbance of articulation, is associated with damage to motor areas of the left hemisphere. By corollary, the right hemisphere, despite a fair capacity for understanding speech, is essentially (that is, apart from a limited capacity for expletive and non-propositional utterance) mute. Interestingly, skilled manual movements (Kimura & Archibald, 1974) and non-verbal oral movements (Mateer & Kimura, 1977) tend also to be impaired in cases of non-fluent aphasia. Moreover, disturbances of sign-language in the deaf are associated with left-hemisphere damage (Kimura, Battison, & Lubert, 1976). After a review of such

evidence. Kimura (1976) suggests that "...the left hemisphere is particularly well adapted, not for symbolic function per se, but for the execution of some categories of motor activity which happened to lend themselves readily to communication" (Kimura, 1976, p. 154).

However, more than motor specialization is involved. Studdert-Kennedy and Shankweiler (1970) concluded, from a study of normal subjects' performance on a test in which competing nonsense syllables were presented simultaneously to left and right ears, that the left hemisphere was specialized for phonological analysis of spoken language. Recent work with split-brain patients has confirmed this conclusion (Zaidel, 1978a). The dissociated right hemisphere of such a patient has a sizeable auditory lexicon and a rudimentary syntax sufficient for understanding phrases of up to three or four words in length. However, it is incapable of identifying nonsense syllables or of recognizing that, say, "rose" rhymes with "toes" (Levy, 1974). In other words, the right hemisphere is not only mute, but is organized by meaning rather than by linguistic structure: Unlike the left hemisphere, it perceives language holistically, seizing meaning from the "auditory contours" of words rather than by phonological analysis. If, as we suggested earlier, the characteristic feature structure of speech sounds derives from articulatory constraints, we should perhaps not be surprised to discover that their perception is linked neurologically to their production.

Direct evidence for a sensorimotor link in the left hemisphere comes from the work of Sussman (1970; Sussman & MacNeilage, 1975; Sussman & Westbury, 1978). Sussman devised a bizarre tracking task in which a sinusoidal waveform, fed into one ear, can be tracked (i.e., copied) by movements of tongue, jaw, lips, or hand. The results of the tracking movements, electronically multiplied into the audio-frequency range, are then fed to the opposite ear. In several experiments, Sussman and his colleagues have shown that tracking movements made by a speech articulator (tongue, jaw, lips) are more accurate if auditory feedback from the movements comes to the right ear (i.e., left hemisphere) rather than to the left. In all but one of the control experiments in which tracking movements were made by hand, there was no ear difference. Sussman and MacNeilage (1975) concluded that their results reflected "a lateralized, speech-related, auditory-sensorimotor integration mechanism" (1975, p. 139).

The ultimate function of such a mechanism is, of course, unknown. However, if anything is to be made of the analogy with bird song, we may speculate that unilateral control is necessary for motor coordination of a bilaterally innervated apparatus (cf. Levy, 1969; Liberman, 1974; Marler, 1970). This might be achieved either by assigning execution primarily to one side of the peripheral apparatus and therefore to lateralized control centers in the brain (as seems to be done in the canary) or by assigning to one side of the brain central coordination of a symmetrically innervated peripheral apparatus (as seems to be done in the human). Lateralization of the associated perceptual center would then follow to facilitate sensorimotor learning. In the human case, evolution of the sensorimotor mechanism led further to development of a lateralized syntactic device, itself perhaps motoric in origin and specialized for precise, temporal coordination of hierarchically ordered structures. The result is that the left hemisphere "...does seem to possess an innate and highly specialized linguistic mechanism

whose paradigmatic functions are phonetic and syntactic encoding and analysis" (Zaidel, 1978a, p. 196).

Finally, in the human case, lateralized control of the vocal apparatus seems to have been laid down on the neural substrate of manual lateralization, already evolved for tool use and/or gestural communication (Levy, 1974). Semmes (1968) has provided an account of the association by arguing (from a lengthy series of gunshot lesions) that the left hemisphere is focally organized for fine motor control, the right hemisphere diffusely organized for broader control. More generally, Zaidel (1978b) has suggested that "...each hemisphere specializes for a different style of information processing..." (p. 263), and Levy (1974) proposes that hemispheric specialization may achieve functional dissociation of neurologically incompatible behaviors. But the important point here is not the possible complementary functions of the cerebral hemispheres (Zangwill, 1960). Rather it is the notion, developed by Kimura (1976) and touched on in our discussion of manual sign language, that the origin of cerebral lateralization for language is in the control of skilled movement rather than in any "higher" symbolic processes.

What is puzzling, of course, is that, unlike song lateralization in birds, which has been observed in virtually every individual studied (Nottebohm, 1977), human lateral specializations are neither uniform across the population nor perfectly associated. The incidence of right-handedness in the U.S. population is estimated at roughly 90% (Hardyck & Petrinovich, 1977; Levy, 1974), and the incidence of left dominance for language at roughly 95% among the right-handed, 60% among the left-handed (Milner, Branch, & Rasmussen, 1964). If such figures prove reliable across the human population, the network of lateralized functions would seem to offer an instance of an "evolutionarily stable strategy" (Maynard-Smith & Price, 1973), a balanced polymorphism that it will be a challenge to explain.

Sensitive Phases

Birds. Many songbirds can only learn their species' song if they are exposed to that song during a sensitive phase. The phase may range from as little as 40 days for the white-crowned sparrow through 10 months for the chaffinch, to as long as two years for the Oregon junco (Petrinovich, 1972). In some birds, such as the white-crowned sparrow or the marsh wren, there may be two distinct phases, separated by weeks or even months: an input phase for perceptual learning and an output phase for subsong and learning to sing. In other birds, such as the chaffinch, the two phases may overlap, with elements of subsong appearing before the input phase has ended. Presumably such variations have adaptive value and can be related to the ecologies and life-histories of the different species. In fact, Immelmann and Suomi (1980) point out that it is precisely the systematic variations across species in temporal patterns of song-learning that validate the concept of a sensitive phase and prove it to be more than a handy descriptive term for a process begun by maturation and ended by song acquisition. Much recent work is therefore aimed at pinning down the ultimate selective pressures (Kroodsma, 1980).

However, the proximate mechanisms controlling onset and offset of sensitive phases are not well understood. Hormone levels are often suggested (e.g., Bateson, 1973). Nottebohm (1967) castrated a male chaffinch during its

first winter, thus precluding either the learning or the singing of song during its first spring. In the second spring the bird was implanted with a testosterone pellet and proved able to learn two tutor songs, but no more. Nottebohm suggests that "...the ability to develop song for the first time is not age-dependent" (1967, p. 278). However, "age" is a cover term for aspects of physical maturity as well as for the mere passage of time. Since castration may have delayed, if not halted, normal maturational processes, the experiment does not rule out physical maturation as the determinant of the onset of song-learning. Since, moreover, a total of two songs falls within the normal chaffinch repertoire range of 2 to 6 songs (Nottebohm, 1967), we might reasonably hypothesize that song-learning had ceased when the available "neural space" was filled (cf. Bateson, in press; Kroodsma, in press). The point here is that, as Immelmann and Suomi (1980) remark, specialized proximate mechanisms beyond physical maturation and neural preemption may not always be necessary for delimitation of a sensitive phase in song birds.

Humans. Lenneberg (1967, pp. 125-187) was the first to postulate a "critical period" for language learning. He was careful to make clear that he was offering no more than an analogy with the critical periods (or sensitive phases) of filial imprinting and song learning in birds. He places the period roughly between the end of the second year and the beginning of the twelfth. Broadly, his argument is based on: (1) the regularity of the time of onset of speech across cultures; (2) the different effects on language of various pathologies, particularly cerebral insult and deafness, as a function of age: in general, the younger the child at the time of brain injury or the older at the time of onset of deafness, the better the prognosis for language development; (3) the commonly observed, increased difficulty of learning a foreign language after puberty—at least without appreciable interference from already known languages. Within the critical period, Lenneberg argued, languages are fully learned by mere exposure; after the critical period they are learned less well and with increasing difficulty—an analogy with song-learning in the zebra finch (Immelmann, 1969).

Lenneberg attributes onset of the "critical period" to general maturation of the central nervous system. Cerebral structure (cell density, dendritic arborization) and chemical composition, as well as characteristic brain wave rhythms measured by electroencephalography, have reached roughly 75 percent of their adult asymptotic values by the age of two years. Thus, Lenneberg does not propose, nor is there any evidence for, a specialized onset mechanism analogous to the changes in hormone levels postulated for some birds.

The lateness of the proposed onset is largely a matter of definition. Since Lenneberg regarded syntax as the distinctive property of language, he identified language onset with the first putting together of words. This typically occurs between 18 and 28 months. Moreover, Lenneberg specifically denied the importance of experience during the first two years, largely on the grounds that children deafened as late as the end of their second year find it no easier to learn language than do those who have been deaf since birth. However, his evidence for this is drawn entirely from informal personal observation, and it seems unlikely that the orderly progression during the first year of life from prespeech oral play through cooing, intonation and babbling is devoid of functional value. If we take the presence of language-specific structure in infant babble at roughly 8 months (Mehler, personal

communication), or even the prespeech lip and tongue movements in train with a mother's behavior (Trevvarthen, Hubley, & Sheeran, 1975), as evidence that language sensitivity has begun, we may place the onset of the sensitive phase in the second half of the first year or even as early as the second month of life. The factors controlling this onset may still then be, as Lenneberg proposed, a combination of physical maturation and appropriate environmental stimulation.

The difficulty of learning a language after puberty is commonly known. Formal evidence for the likelihood of both grammatical and articulatory defects in a second language learned as an adult comes from Oyama (1973, cited by Krashen, 1975). Evidence for even greater defects in a first language learned after puberty has recently come from Genie, a California "wild child" (Curtiss, 1977). When discovered at the age of 13 1/2 years, after nearly twelve years of brutal undernourishment and isolation in a silent back room, Genie had virtually no language. Five years later, she had learned some language by "mere exposure" without specific training. Interestingly, her capacity for phonetic perception was normal, perhaps because her isolation had not begun until 20 months, when the phonetic groundwork had already been laid and she had begun to speak a few words. But her speech was severely distorted and her syntax deficient--for example, she could not use any *wh*-question words, verbal auxiliaries or embedded structures. In other words, she learned language very much less well than a normal child, as Lenneberg would have predicted.

The factors controlling offset at puberty are not known. Lenneberg proposed loss of cerebral plasticity due to completed lateralization of function--without, however, offering any suggestion as to why language should be lateralized. His argument was based on clinical evidence of recovery from aphasia as a function of age. The picture has been confused by recent work suggesting that lateralization may be present from birth (Molfese, Freeman, & Palermo, 1975; Entus, 1977; Glanville, Best, & Levenson, 1977), and essentially complete by five years--roughly coinciding with the time when first language acquisition is approaching completion (Krashen, 1975). But the question of offset mechanism is important if the concept of a sensitive phase for language learning is to retain validity.

The reason for this is that we cannot justify the concept by referring to inter-species differences of the kinds observed in song birds, nor by reference to its onset mechanism, since this appears to correlate with general physical maturation. If, further, its offset mechanism were merely preemption of "neural space," as the articulatory, syntactic and even lexical interference between earlier and later learned languages perhaps suggests, we might be dealing with a general loss in cerebral plasticity and with a process common to other classes of behavior rather than one peculiar to language. In short, the validity of the concept may rest on demonstrating that the offset mechanism is directed specifically at language learning. At present we have no evidence that this is so.

Finally, we must ask what the function of a sensitive phase for language might be (cf. Bateson, in press). First, following Immelmann (1976, p. 152), we must distinguish between the period during which a behavior can be learned and the period during which it normally is learned. It is on the offset of

the former that we might expect selective pressures to bear. If offset were early, roughly contemporaneous with release of offspring into a peer world, the language learned would be that of the parents, and we might reasonably suspect that a sensitive phase ensures a dialect that will attract sexual partners from ecologically similar backgrounds. Dialects might then, indeed, be "signs of incipient speciation" (Marler, 1963, p. 796; cf. Armstrong, 1963, chap. 5). Such a function is unlikely in humans, despite the presumably high correlation between inbreeding and dialect in, say, the highlands of New Guinea or of Austria, because many more salient features (such as habitat and body ornament) serve to isolate human breeding populations.

Moreover, offset in humans is relatively late, well beyond the point where the child has abandoned the nuclear family for its peers. Accordingly, whether a child learns the dialect of its parents rather than of its peers (as is said of some English upper-class children thrown, by the accidents of war, among lower-class peers), or of its peers rather than of its parents (as do the children of non-English-speaking immigrants to Australia or the U.S.A.), may sometimes depend on social rather than directly biological factors. An echo in the behavior of Bewick's wren, which learns the song not of its father but of neighbors in its newly chosen breeding site (Kroodsma, 1974), suggests that social bonding may be among the biological functions of dialects in both bird song and language (cf. Petrinovich, 1972).

Whether this function is important enough to bear the weight of accounting for a sensitive phase in language learning, one may doubt. In fact, given the weakness of this function and the lack of any clear evidence for proximate controlling mechanisms directed specifically at language, one may be tempted to conclude that a "critical period" for human language acquisition is more apparent than real, a mere matter of cerebral maturation in its onset and of neural preemption (or, as in the case of Genie, atrophy) in its offset.

THE INFANT AS PATTERN SEEKER

In songbirds, both species-specific template and sensitive phase are adapted to the same end, namely, acquisition of the species song within a few months of birth. The song to be learned is generally brief and simple. A template ensures that from the varied songs around it, the young bird will learn to recognize (if female) as well as to practice and execute (if male) the song of its own species, while a sensitive phase usually confines learning to the weeks before dispersal from the home site and/or to the months after the bird has settled among its breeding peers. Nonetheless, not all birds that learn to sing have either a template or a sensitive phase. Indeed, certain mimics, such as the North American mockingbird, learn, presumably without template and even late in life, the songs of species quite unrelated to themselves. Perhaps it is among such generalized, all-purpose song learners that we should look for an analogy with the human infant.

In any event, far from being constrained to learn the sound pattern of its language within a few months of birth, the human newborn has before it some two years of infancy. Moreover, what it must learn is not merely to imitate the sounds of the speakers around it--important though this undoubtedly is--but also to perceive and deploy their characteristic sound system. Rather than narrowly defined templates we might therefore expect the infant--

and its caretakers--to have evolved broad behavioral programs that will encourage vocal interchange and facilitate discovery of spoken pattern. The general process of acquisition seems, in fact, to be one of gradual differentiation: sound from silence, voice from sound, mother's voice from stranger's, intonation from monotone, syllabic beat from intonated melody, consonant from vowel, perhaps feature from phoneme.

One-day-old infants will suck a pacifier to turn on music and soon begin to prefer voices to music (Friedlander, 1970). Indeed, within a few days of birth, breast-fed babies have learned to turn toward a voice, twisting the mouth as if in expectation of a nipple and crying when none is there (Alegria & Noirot, Note 1). By 20 to 30 days the infant has learned to recognize its mother's voice, as she reads from behind a screen, and will suck more rapidly for her voice than for a stranger's (Mills & Melhuish, 1974)--provided that she speaks with her customary intonation rather than reads backwards from a text (Mehler, Bertoncini, Barriere, & Jassik-Gerschenfeld, 1978).

From around the second month, the infant becomes accessible to "conversations" with its mother, watching her eyes (humans are the only animals with permanently visible whites to their eyes, contrasting with the iris), smiling, moving lips and tongue in apparent imitation of the mother ("prespeech") and gurgling (Trevvarthen et al., 1975). With the child's discovery that events in the external world--particularly, the vocalizations, touches, gestures of its mother--may be contingent on its own behavior, the way is opened for games (e.g., "peekaboo"), rhythmic interactions, cooing and laughter (Watson, 1977; Papousek & Papousek, 1975). The very precise temporal patterning of mother-infant interaction, with its alternating vocalizations, pauses, exaggerated facial displays, and so on, lays the ground for later social interchange (Stern, Jaffe, Beebe, & Bennett, 1975). Freedle and Lewis (1977) find that vocalization occupies a special place in early mother-infant interaction: It is more likely to accompany playing, looking, holding or touching than changing, feeding or rocking. Moreover, vocalization by one partner is the most likely behavior to follow vocalization by the other, leading to the conclusion that "...vocalization is the central behavior which maintains interaction" (Freedle & Lewis, 1977, p. 160). However, this interactive pattern is not specific to the vocal modality: For deaf children, growing up as signers, signing occupies the privileged position (Feldman et al., 1977). From this we may conclude that mother-infant interaction is broadly adapted to the development, not simply of speech, but of any communicatively viable signaling system. This, in turn, suggests that the infant's discovery of speech may be guided by the pattern of input from its environment rather than by the triggering of tuned detectors.

Of interest here is the nature of the mother's vocalizations, that is, of what has come to be called "baby talk" (BT), the style of speech used by adults, and even young children, when addressing infants (as well as animals and lovers). Baby talk has been studied in many cultures and is characterized by what Ferguson (1978) has termed a "simplified register." The principal acoustic characteristics of this register are, according to Sachs (1978), that it has an overall higher pitch, a wider frequency and intensity range and a more markedly regular rhythmic structure (cf. nursery rhymes). In short, BT exaggerates the acoustic contrasts on which speech is based. While it is unlikely that any single property of the speech addressed to the infant is

essential to normal development (cf. Newport, Gleitman, H., & Gleitman, L. R., 1978), it is equally unlikely that a culturally widespread phenomenon such as BT is devoid of function. If function can be inferred from structure, the function of BT is to draw the infant's attention to important acoustic contrasts in speech (cf. Garnica, 1978) and to launch it on its search for pattern. Thus, we may see BT as the exogenous auditory counterpart of the endogenously controlled eye movements and head turning with which the human newborn searches for visual contour (Haith, 1978).

What the infant has learned perceptually about its native language begins to emerge in babble, around the sixth-to-ninth month. Jakobson (1968) dismisses babble as irrelevant to language acquisition on the grounds that it is primarily a motor activity, devoid of linguistic import. He is correct, inasmuch as normal perception of speech and language, as well as a highly educated level of reading and writing, can be developed, by prolonged and careful instruction, even when articulation has been pathologically precluded since birth (e.g., Fourcin, 1975). But this does not mean that, under normal circumstances, babbling contributes nothing to perceptual or, especially, expressive development. Indeed, it is unlikely that a behavior so regular in its time of onset and developmental course should altogether lack function.

Babble offers an obvious analogy with subsong, the low-intensity, "generalized" singing that precedes true song in many songbirds. Here, too, function is in doubt, because subsong tends to recur each year, as though it might simply reflect lower motivation in early Spring or late Fall (Thorpe, 1956, p. 373). Moreover, the female learns to recognize the male's song even though she herself (like the pathological human cases cited above) never engages in subsong. Nonetheless, subsong does last longer in the bird's first year and bears several interesting analogies with babble--enough to suggest that both activities may be necessary to normal motor, if not to normal perceptual, development.

In the chaffinch, for example, subsong seems to be a poorly differentiated version of the species song with a much greater frequency range. Learning involves dropping unwanted elements and organizing the remaining notes into the correct rhythm (Thorpe, 1956, p. 374), presumably to accord with the inborn template, as modified during early months of the sensitive phase. In the human, babble also seems to begin as a poorly differentiated stream, with many more components than will eventually be used. Gradually, over the course of two or three months, the stream begins to take on properties of the native language, presumably revealing what the infant learned perceptually during its first months of life. Just what these properties are is not yet known, partly because reliable phonetic transcription is difficult. Intonation is the most obvious property, and characteristic pitch contours can be traced in spectrograms (e.g., Nakazima, 1962), but language specific consonant-vowel syllables may be present also (Nakazima, 1975; Kewley-Port & Preston, 1974; Huxley & Ingram, 1971, pp. 162 ff.). In any event, Mehler (personal communication) reports that French-speaking adults can reliably identify infant babble, even in the second month of babbling, as French or not-French.

All this is consistent with the view that babble and subsong enable the organism to discover the limits of its vocal apparatus and to establish necessary sensorimotor links. Here, however, parallels between bird and

infant cease. For while the end of subsong is true song, of which the use does not have to be learned, the end of babble is merely a modest articulatory repertoire, already language-specific, but enough for no more than a start on the discovery of a linguistic system.

The process of discovery is, so far as we know, without parallel in the communication system of any other animal. The infant does not simply imitate, matching a particular utterance to a particular type of situation. Rather, it searches out contrasts among components of its own repertoire and uses them to signal contrasts in its desires, experience or behavior. Often, the contrasts, in both signal and message, are entirely novel and without counterpart in the adult system.

The process is well illustrated in a recent study by Menn (1979). She followed the development of intonation (pitch contour) in the babble and early speech of an American English boy between the ages of about thirteen and fifteen months. She classified his behavioral routines into categories, such as greeting, curiosity, narrative, desiderative, donative. Then she classified the pitch levels of babble in these situations as either moderate or high, and the pitch contours as either rising or falling. Finally, she correlated pitch levels and contours with behavioral routines.

Among the outcomes, predicted from adult speech and observed in the data, were that "narrative" routines were accompanied by falling contours, while "curiosity" or "desiderative" routines were almost always accompanied by rising contours. However, the most interesting finding was that rising "desiderative" contours, addressed to adults, were split according to pitch levels into high (peak above 550 Hz) and moderate (peak below 450 Hz), according to whether the child was seeking an object (e.g., food, toy) or social interaction (e.g., play). In other words, at a stage of his linguistic development when isolated words were still rare and word combinations did not occur at all, this boy had constructed a sub-classification of his own rising pitch contours into "moderate" for sociable occasions and "high" for object-seeking occasions. Since, as Menn (1979) points out, adult speakers of American English do not reliably use absolute pitch to contrast the uses they wish to make of other people, we must conclude that the child had created its own "erroneous" rules of intonation.

Such invention is not without precursor. The process of discovering meaning, and of seeking its correlates in the gestures or vocalizations of others, probably begins with the earliest mother-infant interchanges (cf. MacNamara, 1972; Bruner, 1975). In due course, the infant chances upon such correlates in its own or others' vocal repertoires and, with recognition of the first contrasts in intonation, there begins the slow discovery of sound pattern that will end, several years later, in a full and intricate phonological system. For this and for the parallel processes of syntactic development we find no analogues among birds or apes.

CONCLUSIONS AND QUESTIONS

--A language is an open system, adapted by its dual structure of sound pattern and syntax for unlimited communication. If, as was argued, the dual structure evolved to interface man's intellect with his peripheral anatomy,

it is unlikely that analogous duality of patterning will be found in animals of appreciably lower cognitive complexity.

—A dual structure is also found in manual sign language. That sign languages are manual emphasizes the importance of rapid articulatory gestures to effective linguistic communication. That they also display duality of patterning demonstrates the abstract nature of linguistic capacity: So far as we know, no other animal has developed two essentially equivalent systems of communication using different sensorimotor systems.

—Since none of the supposedly linguistic behaviors of the great apes occurs in a natural environment, recent successes in training them to communicate symbolically have little bearing on the origins of language. However, laboratory studies of the apes may lend insight into the evolution of intelligence and into relations between language and thought.

—Since the capacity for vocal learning has no value beyond its use in communication, its appearance (and pivotal social role) in both man and songbird is of great interest. However, of several possible analogies between birdsong and language learning—auditory templates, sensitive phases and lateralized sensorimotor mechanisms—only the last invites fruitful speculation. Lateralized motor control of birdsong, as well as the association of speech, right-handedness and manual sign-language with left hemisphere mechanisms in humans, suggest that the origin of cerebral lateralization for language may be in the control of skilled sequences of movement. Future work might profitably explore functional relations among manual skills and the perception and production of both speech and sign language, in an attempt to establish the nature and extent of neural overlap.

—The long period of human infancy, taken with the diversity of human languages (both spoken and signed), suggests that biological adaptations for language learning are likely to be behavioral rather than tightly neurophysiological. Study of these behavioral adaptations, particularly of mother-infant interaction during the first year of life, may bring fuller understanding of language and of how it is learned.

REFERENCE NOTE

1. Alegria, J., & Noirot, E. Neonate orientation behavior towards the human voice. Paper read before the XVth International Congress of Ethology, Bielefeld, Germany, August 22-26, 1977.

REFERENCES

- Abbs, J. H., & Sussman, H. M. Neurophysiological feature detectors and speech perception: A discussion of theoretical implications. Journal of Speech and Hearing Research, 1971, 14, 23-26.
- Abercrombie, D. Elements of general phonetics. Chicago: Aldine, 1967.
- Ades, A. E. Adapting the property detectors for speech perception. In R. J. Wales & E. Walker (Eds.), New approaches to language mechanisms. Amsterdam: North-Holland, 1976.
- Armstrong, E. A. A study of bird song. London: Oxford University Press, 1963.

- Bateson, P. P. G. The imprinting of birds. In S. A. Barnett (Ed.), Ethology and development. London: Spastics International Medical Publications (with Heinemann Medical Books Ltd.), 1973.
- Bateson, P. P. G. How do sensitive periods arise and what are they for? Animal Behavior, in press.
- Battison, R. Phonological deletion in American Sign Language. Sign Language Studies, 1974, 5, 1-19.
- Berko, J. The child's learning of English morphology. Word, 1958, 14, 150-177.
- Brown, R. W. A first language: The early stages. Cambridge, Mass.: Harvard University Press, 1973.
- Bruner, J. S. From communication to language--a psychological perspective. Cognition, 1975, 3, 255-287.
- Capranica, R. R. The evoked vocal response of the bullfrog. Cambridge, Mass.: M.I.T. Press, 1965.
- Carney, A. E., Widin, G. P., & Viemeister, N. F. Noncategorical perception of stop consonants differing in VOT. Journal of the Acoustical Society of America, 1977, 62, 961-970.
- Chomsky, N. Three models for the description of language. IRE Transactions on Information Theory, 1956, IT-2, 113-124.
- Chomsky, N., & Halle, M. The sound pattern of English. New York: Harper and Row, 1968.
- Cooper, W. E. Selective adaptation of speech. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. Lindman, & D. B. Pisoni (Eds.), Cognitive theory (Vol. I). Hillsdale, N.J.: Erlbaum Associates, 1975.
- Curtiss, S. Genie: A psycholinguistic study of a modern-day "Wild Child." New York: Academic Press, 1977.
- Dale, P. S. Language development (2nd ed.). New York: Holt, Rinehart & Winston, 1976.
- Dewson, J. H. III. Preliminary evidence of hemisphere asymmetry of auditory function in monkeys. In S. Harnad, R. W. Doty, L. Goldstein, J. Jaynes, & G. Krauthamer (Eds.), Lateralization in the nervous system. New York: Academic Press, 1977, 63-74.
- Du Brul, E. L. Origin of the speech apparatus and its reconstruction in fossils. Brain and Language, 1977, 4, 365-381.
- Eimas, P. D. The relation between identification and discrimination along speech and non-speech continua. Language and Speech, 1963, 6, 206-217.
- Eimas, P. D. Speech perception in early infancy. In L. B. Cohen & P. Salapatek (Eds.), Infant perception. New York: Academic Press, 1975.
- Eimas, P. D., & Corbit, J. D. Selective adaptation of linguistic feature detectors. Cognitive Psychology, 1973, 4, 99-109.
- Eimas, P. D., & Miller, J. L. Effects of selective adaptation on the perception of speech and visual patterns. In R. D. Walk & H. L. Pick, Jr. (Eds.), Perception and experience. New York: Plenum, 1978.
- Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. Speech perception in infants. Science, 1971, 171, 303-306.
- Entus, A. K. Hemispheric asymmetry in processing dichotically presented speech and nonspeech stimuli by infants. In S. J. Segalowitz & F. A. Gruber (Eds.), Language development and neurological theory. New York: Academic Press, 1977, 64-73.
- Feldman, H., Goldin-Meadow, S., & Gleitman, L. Beyond Herodotus: The creation of language by linguistically deprived deaf children. In A. Lock (Ed.), Action, gesture and symbol: The emergence of language.

- New York: Academic Press, 1977.
- Ferguson, C. A. Baby talk as a simplified register. In C. E. Snow & C. A. Ferguson (Eds.), Talking to children. Cambridge: Cambridge University Press, 1978.
- Ferguson, C. A., & Slobin, D. I. Studies of child language development. New York: Holt, Rinehart & Winston, 1973.
- Fourcin, A. J. Language development in the absence of expressive speech. In E. H. Lenneberg & E. Lenneberg (Eds.), Foundations of language development (Vol. 2). New York: Academic Press, 1975, 263-268.
- Freedle, R., & Lewis, M. Prelinguistic conversations. In M. Lewis & L. A. Rosenblum (Eds.), Interaction, conversation and the development of language. New York: Wiley, 1977, 157-186.
- Friedlander, B. A. Receptive language development in infancy. Merrill-Palmer Quarterly, 1970, 16, 7-51.
- Frishberg, N. Arbitrariness and iconicity: Historical change in American Sign Language. Language, 1975, 51, 696-719.
- Frishkopf, L., & Goldstein, M. Responses to acoustic stimuli in the eighth nerve of the bullfrog. Journal of the Acoustical Society of America, 1963, 35, 1219-1228.
- Gardner, R. A., & Gardner, B. T. Teaching sign language to a chimpanzee. Science, 1969, 165, 664-672.
- Gardner, R. A., & Gardner, B. T. Early signs of language in child and chimpanzee. Science, 1975, 187, 752-753.
- Garnica, O. K. Some prosodic and paralinguistic features of speech to young children. In C. E. Snow & C. A. Ferguson (Eds.), Talking to children. Cambridge: Cambridge University Press, 1978, 63-88.
- Glanville, B. B., Best, C. T., & Levenson, R. A cardiac measure of asymmetries in infant auditory perception. Developmental Psychology, 1977, 13, 54-59.
- Haith, M. M. Visual competence in early infancy. In R. Held, H. Leibowitz, & H.-L. Teuber (Eds.), Handbook of sensory physiology (Vol. 8). Berlin: Springer-Verlag, 1978.
- Hardyck, C., & Petrinovich, L. Left-handedness. Psychological Bulletin, 1977, 84, 385-404.
- Hockett, C. F. The origin of speech. Scientific American, 1960, 203, 89-96.
- Huxley, R., & Ingram, E. (Eds.). Language acquisition: Models and methods. New York: Academic Press, 1971.
- Immelmann, K. Song development in the zebra finch and estrildid finches. In R. A. Hinde (Ed.), Bird vocalizations. London: Cambridge University Press, 1969.
- Immelmann, K. Sexual and other long-term aspects of imprinting in birds and other species. In D. S. Lehman, R. A. Hinde, & E. Shaw (Eds.), Advances in the study of behavior (Vol. 4). New York: Academic Press, 1976.
- Immelmann, K., & Suomi, S. Sensitive phases in development. In G. B. Barlow, K. Immelmann, M. Main, & L. Petrinovich (Eds.), Behavior and development. New York: Cambridge University Press, 1980.
- Jakobson, R. Child language, aphasia and phonological universals. The Hague: Mouton, 1968.
- Jakobson, R., Fant, G., & Halle, M. Preliminaries to speech analysis. Cambridge, Mass.: M.I.T. Press, 1963.
- Jenkins, J. J., Jimenez-Pabon, E., Shaw, R. E., & Sefer, J. W. Schuell's aphasia in adults (2nd ed.). New York: Harper and Row, 1975.
- Kelly, D. B., & Nottebohm, F. Projections of a telencephalic auditory

- nucleus--Field L--in the canary. Journal of Comparative Neurology, 1979, 183, 455-470.
- Kewley-Port, D., & Preston, M. Early apical stop production: A voice onset time analysis. Journal of Phonetics, 1974, 2, 195-210.
- Kimura, D. The neural basis of language qua gesture. In H. Whitaker and H. A. Whitaker (Eds.), Studies in neurolinguistics (Vol. 3). New York: Academic Press, 1976.
- Kimura, D., & Archibald, Y. Motor functions of the left hemisphere. Brain, 1974, 97, 337-350.
- Kimura, D., Battison, R., & Lubert, B. Impairment of nonlinguistic hand movements in a deaf aphasic. Brain and Language, 1976, 4, 566-571.
- King, M. C., & Wilson, A. C. Evolution at two levels in humans and chimpanzees. Science, 1975, 188, 107-116.
- Kiparsky, P. Linguistic universals and linguistic change. In E. Bach & R. T. Harms (Eds.), Universals in linguistic theory. New York: Holt, Rinehart & Winston, 1968, 170-202.
- Klima, E. S., & Bellugi, U. The signs of language. Cambridge, Mass.: Harvard University Press, 1979.
- Konishi, M. The role of auditory feedback in the control of vocalization in the white-crowned sparrow. Zeitschrift fur Tierpsychologie, 1965, 22, 770-783.
- Krashen, S. D. The critical period for language acquisition and its possible bases. In D. Aaronson & R. W. Rieber (Eds.), Developmental psycholinguistics and communication disorders. New York: New York Academy of Sciences, 1975, 211-224.
- Kroodma, D. E. Song learning, dialects and dispersal in the Bewick's Wren. Zeitschrift fur Tierpsychologie, 1974, 35, 352-380.
- Kroodma, D. E. Song learning and its sensitive period in the Marsh Wren (*Cistothorus palustris*). In press.
- Kroodma, D. E. Ontogeny of bird song. In G. B. Barlow, K. Immelmann, M. Main, & L. Petrinovich (Eds.), Behavior and development. New York: Cambridge University Press, 1980.
- Kuo, Z.-Y. The dynamics of behavior development (Enlarged ed.). New York: Plenum Press, 1976.
- Ladefoged, P. Preliminaries to linguistic phonetics. Chicago: University of Chicago Press, 1971.
- Lancaster, J. B. Primate communication systems and the emergence of human language. In P. C. Jay (Ed.), Primates. New York: Holt, Rinehart & Winston, 1968.
- Lane, H., Boyes-Braem, P., & Bellugi, U. Preliminaries to a distinctive feature analysis of handshapes in American Sign Language. Cognitive Psychology, 1976, 8, 263-289.
- Lenneberg, E. H. The biological foundations of language. New York: Wiley, 1967.
- Lenneberg, E. H., & Lenneberg, E. (Eds.). Foundations of language development (Vols. 1 and 2). New York: Academic Press, 1975.
- Leppelsack, H. J., & Vogt, M. Responses of auditory neurons in the forebrain of a songbird to stimulation with species-specific sounds. Journal of Comparative Physiology, 1976, 107, 263-274.
- Levi-Strauss, C. The savage mind. Chicago: Chicago University Press, 1968.
- Levy, J. Possible basis for the evolution of lateral specialization of the human brain. Nature, 1969, 224, 614-615.
- Levy, J. Psychobiological implications of bilateral asymmetry. In

- S. J. Dimond & J. G. Beaumont (Eds.), Hemisphere function in the human brain. London: Elek, 1974.
- Lieberman, A. M. The grammars of speech and language. Cognitive Psychology, 1970, 1, 301-323.
- Lieberman, A. M. The specialization of the language hemisphere. In F. O. Schmitt & F. G. Worden (Eds.), The neurosciences: Third study program. Cambridge, Mass.: M.I.T. Press, 1974, 43-56.
- Lieberman, A. M., & Studdert-Kennedy, M. Phonetic perception. In R. Held, H. W. Leibowitz, & H.-L. Teuber (Eds.), Handbook of sensory physiology, Vol. VIII, "Perception." New York: Springer-Verlag, 1978, 143-178.
- Lieberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. Perception of the speech code. Psychological Review, 1967, 74, 431-461.
- Lieberman, P. The speech of primates. The Hague: Mouton, 1972.
- Limber, J. The genesis of complex sentences. In T. Moore (Ed.), Cognitive development and the acquisition of language. New York: Academic Press, 1973.
- Limber, J. Language in child and chimp. American Psychologist, 1977, 32, 280-295.
- MacNamara, J. Cognitive basis of language learning in infants. Psychological Review, 1972, 79, 1-13.
- Marler, P. Inheritance and learning in the development of animal vocalizations. In R-G. Busnel (Ed.), Acoustic behavior of animals. Amsterdam: Elsevier, 1963, 228-243 and 794-797 (addendum).
- Marler, P. Birdsong and speech development: Could there be parallels? American Scientist, 1970, 58, 669-673.
- Marler, P. On the origin of speech from animal sounds. In J. F. Kavanagh & J. E. Cutting (Eds.), The role of speech in language. Cambridge, Mass.: M.I.T. Press, 1975.
- Marler, P., & Peters, S. Selective vocal learning in a sparrow. Science, 1977, 198, 519-521.
- Mateer, C., & Kimura, D. Impairment of nonverbal oral movements in aphasia. Brain and Language, 1977, 4, 262-276.
- Mattingly, I. G. The human aspects of speech. In J. F. Kavanagh & J. E. Cutting (Eds.), The role of speech in language. Cambridge, Mass.: M.I.T. Press, 1975, 63-72.
- Maynard-Smith, J., & Price, G. R. The logic of animal conflict. Nature, 1973, 246, 15.
- Mayr, E. Behavior programs and evolutionary strategies. American Scientist, 1974, 62, 650-659.
- Mehler, J., Bertoncini, J., Barriere, M., & Jassik-Gershenfeld, D. Infant recognition of mother's voice. Perception, 1978, 7, 491-497.
- Menn, L. Pattern, control and contrast in beginning speech. Bloomington: Indiana University Linguistics Club, 1979.
- Miller, D. B. Conceptual strategies in behavioral development: Normal development and plasticity. In G. B. Barlow, K. Immelmann, M. Main, & L. Petrinovich (Eds.), Behavior and development. New York: Cambridge University Press, 1980.
- Miller, G. A., Galanter, E., & Pribram, K. Plans for the structure of behavior. New York: Holt, Rinehart & Winston, 1960.
- Molfese, D. L., Freeman, R. B., & Palermo, D. S. The ontogeny of brain lateralization for speech and non-speech stimuli. Brain and Language, 1975, 2, 356-368.
- Mills, M., & Melhuish, E. Recognition of mother's voice in early infancy.

- Nature, 1974, 252, 123.
- Milner, B., Branch, C., & Rasmussen, T. Observations on cerebral dominance. In A. V. S. deRueck & M. O'Connor (Eds.), Disorders of language. London: Churchill, 1964, 200-214.
- Nakazima, S. A comparative study of the speech developments of Japanese and American English in childhood (1)--A comparison of the developments of voices at the prelinguistic period. Studia Phonologica, 1962, 2, 27-46.
- Nakazima, S. Phonemicization and symbolization in language development. In E. H. Lenneberg & E. Lenneberg (Eds.), Foundations of language development (Vol. 1). New York: Academic Press, 1975, 181-187.
- Neuweiler, G. Recognition mechanisms in echolocation of bats. In T. H. Bullock (Ed.), Recognition of complex acoustic stimuli. Berlin: Dahlem Konferenzen, 1977, 111-126.
- Newport, E. L., Gleitman, H., & Gleitman, L. R. Mother, I'd rather do it myself: Some effects and non-effects of maternal speech style. In C. E. Snow & C. A. Ferguson (Eds.), Talking to children. Cambridge: Cambridge University Press, 1978.
- Nottebohm, F. The role of sensory feedback in the development of avian vocalizations. Proceedings of the XIVth International Ornithological Conference, Oxford, 1967, 265-280.
- Nottebohm, F. Ontogeny of bird song. Science, 1970, 167, 950-956.
- Nottebohm, F. Neural lateralization of vocal control in a passerine bird. I. Song. Journal of Experimental Zoology, 1971, 177, 229-261.
- Nottebohm, F. Neural lateralization of vocal control in a passerine bird. II. Subsong, calls and a theory of vocal learning. Journal of Experimental Zoology, 1972, 179, 35-49.
- Nottebohm, F. A zoologist's view of some language phenomena with particular emphasis on vocal learning. In E. H. Lenneberg & E. Lenneberg (Eds.), Foundations of language development: A multi-disciplinary approach (Vol. 1). New York: Academic Press, 1975.
- Nottebohm, F. Phonation in the orange-winged Amazon parrot, (Amazona amazonica). Journal of Comparative Physiology, 1976, 108, 157-170.
- Nottebohm, F. Asymmetries in neural control of vocalization in the canary. In S. Harnad, R. W. Doty, L. Goldstein, J. Jaynes, & G. Krauthamer (Eds.), Lateralization in the nervous system. New York: Academic Press, 1977, 23-44.
- Nottebohm, F., & Nottebohm, M. E. Left hypoglossal dominance in the control of canary and white-crowned sparrow song. Journal of Comparative Physiology, 1976, 108, 171-192.
- Oyama, S. A sensitive period for the acquisition of a second language. Unpublished doctoral dissertation, Harvard University, 1973.
- Papousek, H., & Papousek, M. Cognitive aspects of preverbal social interaction between human infants and adults. In M. O'Connor (Ed.), Parent-infant interaction. Amsterdam: Elsevier, 1975, 241-260.
- Pastore, R. E., Ahroon, W. A., Baffuto, K. J., Friedman, C., Ruele, J. S., & Fink, E. A. Common factor model of categorical perception. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 686-696.
- Patterson, F. G. The gestures of a gorilla: Language acquisition in another pongid. Brain and Language, 1978, 5, 72-97.
- Penfield, W., & Roberts, L. Speech and brain mechanisms. Princeton, N.J.: Princeton University Press, 1959.
- Petersen, M. R., Beecher, M. D., Zoloth, S. R., Moody, D. B., & Stebbins,

- W. C. Neural lateralization of species-specific vocalizations by Japanese macaques (*Macaca fuscata*). Science, 1978, 202, 324-327.
- Petrinovich, L. Psychobiological mechanisms in language development. In G. Newton & A. H. Riesen (Eds.), Advances in psychobiology (Vol. 1). New York: Wiley, 1972.
- Pisoni, D. B. Identification and discrimination of the relative onset of two component tones: Implications for the perception of voicing in stops. Journal of the Acoustical Society of America, 1977, 61, 1352-1361.
- Premack, D. Intelligence in ape and man. Hillsdale, N.J.: Erlbaum, 1976.
- Premack, D., & Woodruff, G. Chimpanzee problem-solving: A test for comprehension. Science, 1978, 202, 532-535.
- Remez, R. E. Adaptation of the category boundary between speech and non-speech: A case against feature detectors. Cognitive Psychology, 1979, 11, 38-57.
- Rosen, S. M. Range and frequency effects in consonant categorization. Journal of the Acoustical Society of America, in press.
- Rumbaugh, D. Language learning by a chimpanzee. New York: Academic Press, 1977.
- Sachs, J. The adaptive significance of linguistic input to prelinguistic infants. In C. E. Snow & C. A. Ferguson (Eds.), Talking to children. Cambridge: Cambridge University Press, 1978, 51-62.
- Semmes, J. Hemispheric specialization: A possible clue to mechanism. Neuropsychologia, 1968, 6, 11-26.
- Simon, H. J., & Studdert-Kennedy, M. Selective anchoring and adaptation of phonetic and nonphonetic continua. Journal of the Acoustical Society of America, 1978, 64, 1338-1357.
- Siple, P. (Ed.) Understanding language through sign language research. New York: Academic Press, 1979.
- Stern, D. N., Jaffe, J., Beebe, B., & Bennett, S. L. Vocalizing in union and in alternation: Two modes of communication within the mother-infant dyad. In D. Aaronson & R. W. Rieber (Eds.), Developmental psycholinguistics and communication disorders. New York: New York Academy of Sciences, 1975, 89-100.
- Stetson, R. H. Motor phonetics. Amsterdam: North Holland, 1952.
- Stevens, K. N. The potential role of property detectors in the perception of consonants. In G. Fant & M. A. A. Tatham (Eds.), Auditory analysis and perception of speech. New York: Academic Press, 1975, 303-330.
- Stevens, K. N., & Halle, M. Remarks on analysis by synthesis and distinctive features. In W. Wathen-Dunn (Ed.), Models for the perception of speech and visual form. Cambridge, Mass.: M.I.T. Press, 1967.
- Stokoe, W. C., Jr. Classification and description of sign languages. In T. A. Sebeok (Ed.), Current trends in linguistics (Vol. 12). The Hague: Mouton, 1974, 345-371.
- Stokoe, W. C., Jr., Casterline, D. C., & Croneberg, C. G. A dictionary of American Sign Language on linguistic principles. Washington, D.C.: Gallaudet College Press, 1965.
- Strange, W., & Jenkins, J. J. The role of linguistic experience in the perception of speech. In H. L. Pick, Jr. & R. D. Walk (Eds.), Perception and experience. New York: Plenum, 1978.
- Studdert-Kennedy, M. The perception of speech. In T. A. Sebeok (Ed.), Current trends in linguistics (Vol. 12). The Hague: Mouton, 1974, 2349-2385.
- Studdert-Kennedy, M. Speech perception. In N. J. Lass (Ed.), Contemporary

- issues in experimental phonetics. New York: Academic Press, 1976.
- Studdert-Kennedy, M. Universals in phonetic structure and their role in linguistic communication. In T. H. Bullock (Ed.), Recognition of complex acoustic signals. Berlin: Dahlem Konferenzen, 1977.
- Studdert-Kennedy, M., & Shankweiler, D. P. Hemispheric specialization for speech perception. Journal of the Acoustical Society of America, 1970, 48, 579-594.
- Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., & Cooper, F. S. The motor theory of speech perception: A reply to Lane's critical review. Psychological Review, 1970, 77, 234-249.
- Sussman, H. M. A laterality effect in lingual auditory tracking. Journal of the Acoustical Society of America, 1970, 49, 1874-1880.
- Sussman, H. M., & MacNeilage, P. F. Studies of hemispheric specialization for speech production. Brain and Language, 1975, 2, 131-151.
- Sussman, H. M., & Westbury, J. R. A laterality effect in isometric and isotonic labial tracking. Journal of Speech and Hearing Research, 1978, 21, 563-579.
- Terrace, H. S., Petitto, L., & Bever, T. G. Project Nim: Progress report I. New York: Department of Psychology, Columbia University, 1976. (a)
- Terrace, H. S., Petitto, L., & Bever, T. G. Project Nim: Progress report II. New York: Department of Psychology, Columbia University, 1976. (b)
- Thorpe, W. H. Learning and instinct in animals. Cambridge, Mass.: Harvard University Press, 1956.
- Trevarthen, C., Hubley, P., & Sheeran, L. Les activites innées du nourrisson. La Recherche, 1975, 56, 447-458.
- Umiker-Sebeok, D. J., & Sebeok, T. A. Aboriginal sign 'Languages' from a semiotic point of view. Ars Semeiotica, 1977, 1, 69-97.
- van Lawick-Goodall, J. In the shadow of man. London: Collins, 1971.
- Watson, J. S. Perception of contingency as a determinant of social responsiveness. In E. B. Thomas (Ed.), The origins of the infant's social responsiveness. Hillsdale, N.J.: Erlbaum, 1977.
- West, LaMont, Jr. The sign language: An analysis. Unpublished doctoral dissertation, Indiana University, Bloomington, 1960.
- Whitaker, H., & Whitaker, H. A. (Eds.). Studies in neurolinguistics (Vols. 1-3). New York: Academic Press, 1976.
- Whitfield, I. C., & Evans, E. F. Responses of auditory cortical neurons to stimuli of changing frequency. Journal of Neurophysiology, 1965, 28, 655-672.
- Wilson, E. O. Sociobiology: The new synthesis. Cambridge: The Belknap Press, 1975.
- Wollberg, Z., & Newman, J. D. Auditory cortex of squirrel monkey: Response patterns of single cells to species-specific vocalizations. Science, 1972, 175, 212-214.
- Youngren, O. M., Peek, F. W., & Phillips, R. E. Repetitive vocalizations evoked by local electrical stimulation of avian brains. Brain, Behavior and Evolution, 1974, 9, 393-421.
- Zaidel, E. Lexical organization in the right hemisphere. In P. A. Buser and A. Rougeul-Buser (Eds.), Cerebral correlates of conscious experience. Amsterdam: Elsevier/North-Holland Biomedical Press, 1978, 177-197. (a)
- Zaidel, E. Concepts of cerebral dominance in the split brain. In P. A. Buser and A. Rougeul-Buser (Eds.), Cerebral correlates of conscious experience. Amsterdam: Elsevier/North-Holland Biomedical Press, 1978, 263-285. (b)
- Zangwill, O. Cerebral dominance and its relation to psychological function. Edinburgh: Oliver & Boyd, 1960.

PROOFREADING ERRORS ON THE WORD the: NEW EVIDENCE ON READING UNITS*

Alice F. Healy*

Abstract. In three experiments, subjects read passages and circled misspellings in them. In Experiment 1, misspellings were introduced by transposing two adjacent letters in a word. Subjects made a disproportionately small number of errors on the word the in the transposition-proofreading task. In Experiments 2 (prose passages) and 3 (scrambled nouns), misspellings were introduced by replacing instances of the letter t with the letter z. Letter-detection tasks in which subjects searched for instances of t in passages without misspellings were compared to the substitution-proofreading tasks in which the subjects, in effect, searched for z. Subjects in Experiment 2 made a disproportionately large number of errors on the in letter detection but not in proofreading. In Experiment 3, subjects made more errors on common than on rare nouns in letter detection but not in the proofreading. The results provide evidence that common words are normally read in units larger than letters but are read in letter units when they are misspelled.

INTRODUCTION

Letter-detection tasks have been used to obtain evidence for the size of the units used in reading (Healy, 1976; Drewnowski & Healy, 1977). The present study sought to obtain evidence on the same issue by employing various proofreading tasks and by comparing proofreading and detection tasks.

Subjects have been found to be especially likely to make errors on the word the in letter-detection tasks (Corcoran, 1966; Healy, 1976). By ruling out hypotheses concerning the pronunciation and location of the target letters and the semantic and syntactic redundancy of the, Healy (1976) argued that the preponderance of letter-detection errors on the was due to its high frequency, which made it especially likely to be read as a unit, or chunk, rather than in

*This paper is to appear in the Journal of Experimental Psychology: Human Perception and Performance, 6.

*Also Yale University.

Acknowledgment: This research was supported in part by PHS Grants MH26573 and RR07015, NSF Grant BNS 77-00077, and a grant from The Spencer Foundation to Yale University and NICHD Grant HD01994 and BRS Grant RR05596 to Haskins Laboratories. The author was supported by a Junior Faculty Fellowship from Yale University during the preparation of this manuscript. The author is indebted to C. Paul for help with the design and conduct of the experiments, to J. Gottlieb and J. Chung for help with the data analyses, and to R. G. Crowder, A. Drewnowski, W. K. Estes, M. Kubovy, A. G. Levitt, and B. H. Repp for helpful discussions about this research.

terms of its component letters. As further support for this argument, Healy (1976) demonstrated that in a passage of scrambled nouns, subjects were more likely to make letter-detection errors on common nouns than on rare nouns. These findings were extended by Drewnowski and Healy (1977) who employed the trigram the as well as the letter t as a detection target. A preponderance of detection errors was found on the word the, rather than on words with embedded the trigrams (such as bathed), for both targets. In addition, more detection errors were found on the word the when it occurred in an appropriate syntactic context than when it did not. Finally, more errors occurred on the word the in a passage typed in standard paragraph format than in passages in which word-group identification was disturbed by the use of mixed typecases or a list format. On the basis of these findings, Drewnowski and Healy (1977) concluded that familiar word sequences may be read in units larger than the word, probably short syntactic phrases or word frames, such as "on the ---."

A set of five hypotheses about the reading process that are consistent with the findings from detection tasks was proposed by Drewnowski and Healy (1977). These hypotheses will be referred to here as the "unitization hypotheses." Specifically, (1) a hierarchy of processing levels was identified and was defined in terms of the units available at each level, such as letters, words, and phrases. (2) It was proposed that the completion of processing at a given level is tantamount to the identification of the unit at that level. In accordance with this hypothesis, detection tasks that require subjects to identify targets at a given level allow one to monitor the completion of processing at that level. For example, letter-detection tasks allow one to monitor the completion of processing at the letter level. It was further postulated (3) that subjects process text in parallel at the various levels available to them and (4) that once a unit has been identified at a given level, the subjects will proceed to the next unit at that level without necessarily completing the processing of units at the lower levels in the hierarchy. For example, once the subjects have identified a word, they may move on to the next word in the text, which they will process at all levels in parallel, without necessarily completing the processing of all the letters within the word just identified. (5) Familiarity with a unit at a given level should facilitate processing of it. For example, common words should be processed at the word level more easily than rare words. In particular, the, which is the most common word in English, should be processed at the word level more easily than other words.

It should be noted that the unit processing depicted by these hypotheses is what has been described elsewhere as "automatic processing" (see, for example, LaBerge & Samuels, 1974; Shiffrin & Schneider, 1977). Hence, it is assumed that when a given unit is processed, the constituent elements of that unit may be hidden from conscious perception. For example, when a word unit is processed, the constituent letters of that word may be hidden from conscious perception, even when the word itself is consciously perceived. The hypothesis that familiarity of a given unit facilitates processing is compatible with the extensive evidence that automatic processes require considerable training to develop. Despite the fact that these hypotheses describe automatic processes, rather than the slower controlled processes, the hypotheses are certainly compatible with the possibility that controlled processes may be occurring in the reading task as well as automatic processes. For example, in the case when all the letters in a given word are identified before the word itself, the subjects may use controlled processing to identify

the word before moving on to the next string of letters. In fact, such a strategy seems reasonable for subjects who are reading for meaning. Although it seems reasonable to postulate controlled processing of a higher-level unit after the automatic processing of the constituent lower-level units is completed, it seems less reasonable to postulate controlled processing of the constituent lower-level units after the automatic processing of the higher-level unit is completed. Indeed such an asymmetry between the levels of processing is implied by the fourth hypothesis.

On the basis of the unitization hypotheses, one would expect that subjects would be likely to make many letter-detection errors on the word the when reading standard text, since they would tend to complete processing at the word or phrase level before the letter level. More generally, these hypotheses lead one to expect more letter-detection errors on common words than on rare words, since the probability of faster processing at the word level than at the letter level would be greater for common words than for rare words. Another prediction from these hypotheses is that the tendency to make letter-detection errors on the would be greatly reduced if the processing of units larger than the letter were disturbed by typing every other letter in capitals. These predictions have all been confirmed in the previous studies by Healy (1976) and Drewnowski and Healy (1977) and were examined again in the present study, along with an examination of proofreading errors.

What do the unitization hypotheses lead one to expect about errors subjects make in proofreading? In order to answer this question, we must be more explicit about the nature of the proofreading task. In the proofreading task used in the first experiment of the present study, which we call a "transposition-proofreading" task, subjects were told to encircle every instance of a misspelling, and misspellings were introduced into a prose passage by transposing two adjacent letters within a word. Clearly, in such a task the processing of the misspelled words should be minimally disturbed at the letter level, since they contain the same letters as their correct versions. Effects of letter transpositions would be expected at the letter level only to the extent that there are sequential dependencies in letter recognition. (For example, processing of the letter u may be facilitated if it occurs after the letter q.) However, automatic processing at the word level and above should be greatly disturbed, if not prohibited, for the misspelled words, since no word-level unit would correspond to the anomalous sequence of letters. Subjects would be limited to the slower controlled processing at the word level and above. On the basis of these considerations alone, one would expect no difference between the and other words. However, an additional factor is also relevant to this task--namely the subjects' ability to identify a letter string as a misspelling. Subjects should find it easier to identify a misspelling caused by a single transposition in a very short word such as the than in a longer word, because in a short word the misspelled letter string would have less in common with the correctly spelled word. (See Holbrook, 1978a, for a similar argument and a demonstration that the perceived similarity between a word in its misspelled and correct forms depends on word length.) Thus, according to these hypotheses, in this transposition-proofreading task subjects should make relatively few errors on the, as opposed to the preponderance of errors on the in letter-detection tasks.

Although the unitization hypotheses lead one to expect large differences in the patterns of errors in the transposition-proofreading task and in a comparable letter-detection task, other reasonable views would lead one to expect similar patterns in the two tasks. For example, Corcoran (1966) explained the preponderance of letter-detection errors on the word the in terms of the redundancy of the. Healy (1976) termed this the "redundancy hypothesis." Although there are several different kinds of redundancy (Smith, 1971), only semantic and syntactic redundancy is referred to by this hypothesis. Specifically, Corcoran (1966) suggested that the may be "'taken for granted' and thus not scanned" (p. 658). (See Hatch, Polin, & Part, 1974, for an expanded discussion of the importance of predictability, or syntactic and semantic redundancy, in this task.) According to this hypothesis, a preponderance of errors would be expected on the in the transposition-proofreading task as well as in a comparable letter-detection task, since the would be syntactically and semantically redundant in both situations. If subjects fail to scan the because they take it for granted, they would not be able to detect misspellings of it.

Similarly, Schindler (1978) proposed an explanation in terms of eye-movement patterns for his finding that subjects made more letter-detection errors on function words than on content words. He described results from an experiment by Rayner (1977) demonstrating that the received fewer and usually shorter fixations than other words. Although Schindler did not propose an explanation for the eye-movement patterns, it is likely that they would be guided by the subjects' expectations based on prior word context. Words expected to be of little informational content would be likely to be skipped. Schindler (1978) also considered the possibility that subjects give very little visual attention to words that are likely to be unimportant. Both of these hypotheses by Schindler are consistent with, if not merely restatements of, the redundancy hypothesis.

A recent proofreading experiment by Holbrook (1978b) provides some direct support for the redundancy hypothesis. Holbrook found a significant positive correlation between the subjective verbal uncertainty of words, as measured by a Cloze test (Taylor, 1953), and the detection of typographical errors in those words.

The results of an earlier proofreading task, although inconclusive, might also lead one to expect a similar pattern of results in the transposition-proofreading and comparable letter-detection tasks. Corcoran (1967) conducted a proofreading experiment in which letters were omitted from various words in a prose passage, and subjects were asked to indicate the locations where letters were missing. As in his earlier letter-detection task (Corcoran, 1966), Corcoran found that subjects made more errors on silent es than on pronounced es in the proofreading task. Although errors were most frequent on the when the task was to detect the letter e (Corcoran, 1966), when the task was proofreading, the probability of failing to detect a missing e from the was roughly equivalent to the overall probability of failing to detect a missing letter. Corcoran (1967) does indicate, though, that subjects made significantly more errors in proofreading on the e in the than on other terminal pronounced es. The results from Corcoran's (1967) proofreading task are, therefore, somewhat ambiguous for the. It is also difficult to determine what to expect on the basis of the unitization hypotheses in proofreading with

omitted letters, because automatic processing at both the word and letter levels (in the case of the omitted letters) would be virtually prohibited. For that reason Corcoran's proofreading task was not employed in the present study.

EXPERIMENT 1

Method

Subjects. Ninety-six male and female Yale undergraduates participated as subjects. The data from five additional subjects were not analyzed because those subjects had participated previously in similar experiments.

Design and materials. A single typewritten passage was employed. The passage was based on a 321-word prose passage taken from The Social Animal by Elliot Aronson. Forty misspellings were introduced into this passage in a pseudorandom fashion so that misspellings occurred on exactly two words in every block of 16, excluding the final word of the passage. Each of these misspellings involved a transposition of two adjacent letters in a word. None of these transpositions yielded a new word except for two which yielded very infrequent words (fro from for and eh from he)¹. Exactly 11 of the 40 misspellings involved the word the. (There were 38 thes in the passage as a whole.) Six of the misspellings of the were obtained by transposing the last two letters, forming a letter string which is pronounceable (teh), and five of the misspellings were obtained by transposing the first two letters, forming a letter string which is not pronounceable (hte).

Each subject was shown a mimeographed copy of the passage, preceded by a mimeographed sheet of instructions.

Procedure. The subjects were tested in a group session conducted in a classroom. The subjects were instructed to read the prose passage at their "normal reading speed," but whenever they came to a spelling error they were to encircle it with their pen or pencil. The subjects were told that if at any time they realized that they missed an error in a previous word, they should not retrace their steps to encircle it and that they should not slow down their reading speed in order to be overcautious about getting the errors.

Results and Discussion

The results are summarized in Table 1, which includes the means and the standard errors of the means for the percentage of proofreading errors made by the subjects (out of 40 possible errors), for the percentage of proofreading errors on the (out of 11 possible errors), and for the conditional percentage of proofreading errors on the given a proofreading error. All errors considered here and in the subsequent analyses in this paper were omission errors (misses). The conditional percentage of proofreading errors on the was derived for a given subject by determining the ratio of the number of proofreading errors on the to the total number of proofreading errors. By chance alone, the conditional percentage should be 27.5%, since 11 of the 40 transpositions occurred in the. Healy (1976) and Drewnowski and Healy (1977) found this measure to be the most sensitive index of performance in their

detection tasks, since it is unaffected by the speed-accuracy tradeoff typically found in such tasks.

Table 1

Means and Standard Errors of Means for Percentages of Transposition-Proofreading Errors in Experiment 1

Error percentage	$p(\text{error})$	$p(\text{error on the})$	$p(\text{error on the} \text{error})$ percentage ^a	N'
M	11.6	3.3	6.2	93
SE _M	0.8	0.7	1.4	

Note. In this table and in the succeeding tables in this paper, the total number of subjects ($N = 96$) does not equal the number of subjects on which the mean conditional error percentage is based (N'), since not all subjects made errors on the passage.

^aThe value of $p(\text{error on the}|\text{error})$ expected by chance alone is 27.5.

Although subjects failed to detect about 12% of all misspellings, they missed only about 3% of the misspellings that involved the. In fact, of the 93 subjects who made proofreading errors on this passage, only 26 made any error on the. The conditional percentage of proofreading errors on the given an error was significantly below chance level, $t(92) = 21.5$, $p < .001$. Contrary to the redundancy hypothesis, it is clear from these results that subjects do not fail to scan the when proofreading, but rather are extremely accurate at detecting misspellings of this word. Further, the pronounceability of misspellings seems to be of little consequence, since very few errors were made on the, whether it was misspelled as feh or as hte.

In order to test directly the hypothesis that the subjects' ability to identify a letter string as a misspelling depends on word length, the percentage of proofreading errors was computed as a function of word length for all misspelled words excluding the. (See Table 2.) There were six misspelled two-letter words in the passage, four misspelled three-letter words (excluding the), six misspelled four-letter words, four misspelled five-letter words, and nine misspelled words six to ten letters long. In accordance with the hypothesis proposed above, subjects made more errors on the words five to ten letters long than on the shorter words, $F(4,380) = 81.4$, $MS_e = 162$, $p < .001^2$. Furthermore, the percentage of errors on the (see Table 1) was slightly, but significantly, greater than the percentage of errors on other three-letter words (see Table 2), $F(1,95) = 14.3$, $MS_e = 26$, $p < .001^3$.

Table 2

Means and Standard Errors of Means for Percentages
of Transposition-Proofreading Errors Excluding those on the word the
in Experiment 1 as a Function of Word Length

Error percentage	Word length				
	2	3	4	5	6-10
M	7.6	0.5	6.2	22.9	27.9
SE _M	1.2	0.5	1.1	1.8	2.1

EXPERIMENT 2

Although the results of Experiment 1 are compatible with the unitization hypotheses, the additional hypothesis concerning the subjects' ability to identify a letter string as a misspelling must be added to account for performance. (On the basis of the unitization hypotheses alone, no differences between misspellings of the and of other words were expected, since no word-level units would correspond to the misspelled letter strings. However, the hypothesis that subjects would find it easier to identify a misspelling in a short word than in a longer word correctly predicted that subjects would make relatively few errors on the short word the.) For this reason, we constructed a new "substitution-proofreading" task that eliminated the need for the subjects to identify a letter string as a misspelling. The new task therefore enabled us to test more directly the unitization hypotheses as they apply to proofreading. The new task also had the advantage of permitting an elegant comparison of proofreading and letter detection.

Specifically, subjects were told to encircle every instance of a misspelling, and misspellings were introduced by replacing each instance of the letter t with the letter z. Subjects were informed of this fact as well as of the important fact that there were no other zs in the passage, so that each z represented a misspelling. Superficially, this substitution-proofreading task is strictly analogous to a detection task in which subjects search for ts in the corresponding passage without misspellings. According to the unitization hypotheses, though, the pattern of results should be very different for the analogous substitution-proofreading (z-circling) and letter-detection (t-circling) tasks. In the substitution-proofreading task, as in the transposition-proofreading task, processing at the letter level should be minimally affected in the misspelled words. Such an effect is expected only to the extent that the subjects' baseline ability to detect z is different from their baseline ability to detect t. In addition, as in the transposition-proofreading task, automatic processing of a given letter sequence at the word level and above should be greatly disturbed, if not

prohibited, by z substitutions. Subjects would be restricted to the slower controlled processing at these higher levels. The additional factor considered for the transposition-proofreading task--that the subjects' ability to identify a letter string as a misspelling depends on word length--should not be relevant to the present substitution-proofreading task, since it is made clear to the subjects that all zs represent misspellings. Considering the factors that are relevant to this task, one would not expect subjects to make either a disproportionately large or a disproportionately small number of errors on the in the substitution-proofreading task, but one would expect subjects to make a disproportionately large number of errors on the in the analogous letter-detection task.

Although a different pattern of results is expected for the present substitution-proofreading and letter-detection tasks according to the unitization hypotheses, no difference between the two tasks is expected according to the redundancy hypothesis, because the syntactic and semantic redundancy of the would not be changed by substituting z for t.

The effect of mixed typecase was also examined in this experiment. As in the study by Drewnowski and Healy (1977, Experiment 3), subjects were given passages typed in standard format and passages typed with every other letter in capitals in order to disturb the use of reading units larger than the letter. According to the unitization hypotheses, the conditional percentage of errors on the given an error should be reduced in the passages with mixed typecase relative to the passages typed in the standard fashion. In contrast, the redundancy hypothesis could not provide a simple account of any differences in conditional percentages for the two types of passages, since the syntactic and semantic redundancy of the would not change with a change in typecase.

Method

Subjects. The same subjects were employed as in Experiment 1. The subjects performed Experiment 2 immediately after completing Experiment 1.

Design and materials. Four passages were used, all based on a 100-word prose passage from Golding's Lord of the Flies. The basic passage (the "unmixed t passage") included 40 ts, 11 of them in the word the. This passage was identical to that employed by Healy (1976, Experiment 1). The second passage ("unmixed z passage") was identical to the unmixed t passage except that every t was replaced by a z. There were no other zs. None of the letter strings containing a z in the unmixed z passage were English words. The third and fourth passages ("mixed t and mixed z passages") were identical to the unmixed t and unmixed z passages, respectively, except that every other letter was typed in capitals. There were two versions of these passages. In version A the odd letters were capital, and in version B, the even letters were capital.

Each subject was shown mimeographed copies of all four passages, typed on separate sheets of paper and stapled together. For all subjects, the four passages were divided into two sets--the t passages and the z passages. Each set of passages was preceded by a mimeographed sheet of instructions. The order of the two sets was counterbalanced across subjects. A given subject

was shown only one version (A or B) of the mixed passages. The version of the mixed t passage shown to a given subject matched the version of the mixed z passage shown to him or her (e.g., subjects shown version A of the mixed t passage were shown version A of the mixed z passage). Furthermore, the order of the unmixed and mixed passages was counterbalanced across subjects, but the order of the unmixed and mixed t passages shown to a given subject corresponded to the order of the unmixed and mixed z passages shown to him or her. The three divisions among subjects (t first versus z first; version A versus version B of mixed passages, and order of unmixed and mixed passages) were orthogonal to each other, so that there were approximately equal numbers of subjects (11-13) in the eight subgroups of subjects.

Procedure. The subjects performed the experiment in two stages, one stage for each set of passages. In a given stage the subjects read the instructions for the appropriate set of passages, were given an opportunity to ask questions about the instructions, and then were allowed to perform the task for that set of passages.

The subjects were instructed to read the t passages at their "normal reading speed," but whenever they came to a letter t, they were to encircle it with their pen or pencil. In analogy with the instructions for Experiment 1, the subjects were told that if at any time they realized that they missed a t in a previous word, they should not retrace their steps to encircle it and that they should not slow down their reading speed in order to be overcautious about getting the ts. These instructions were identical to those used by Healy (1976).

The subjects were told that the z passages would each contain a number of spelling errors all of the same type: Each instance of the letter t was replaced by the letter z. They were further told that there were no other zs in the passages. The other instructions for the z passages were analogous to those for the t passages except the letter t was replaced by z.

Results and Discussion

The results are summarized in Table 3, which is analogous to Table 1 except that it includes data from two tasks (detection and substitution proofreading) for two passages (mixed and unmixed). The percentage of errors was greater in the detection (t-circling) task (mean = 15%) than in the substitution-proofreading (z-circling) task (mean = 5%), $F(1,95) = 91.0$, $MS_e = 114$, $p < .001$. Overall, the percentage of errors made in the unmixed passage (mean = 12%) was greater than in the mixed passage (mean = 9%), $F(1,95) = 10.8$, $MS_e = 85$, $p = .002$, and, whereas there was a large difference in error percentages between the two passages for t detection, the difference was smaller and in the opposite direction for proofreading, $F(1,95) = 65.9$, $MS_e = 51$, $p < .001$. A similar pattern of results was found when the percentage of errors on the was considered. The percentage of errors on the in t detection (mean = 25%) was greater than in proofreading (mean = 4%), $F(1,95) = 103.7$, $MS_e = 453$, $p < .001$. Overall the percentage of errors on the in the unmixed passage (mean = 21%) was greater than in the mixed passage (mean = 8%), $F(1,95) = 81.2$, $MS_e = 175$, $p < .001$, and whereas the difference between the two passages in percentages of errors on the was large for t detection, it was smaller and in the opposite direction for proofreading, $F(1,95) = 103.8$, $MS_e = 146$, $p < .001$.

Table 3

Means and Standard Errors of Means (in Parentheses) for
Letter-Detection (t-circling) and Substitution-Proofreading (z-circling)
Error Percentages in Experiment 2
as a Function of Passage Type

Task	p(error)	p(error on <u>the</u>)	p(error on <u>the</u> error) percentage ^a	N ¹
Detection				
Unmixed	19.7 (1.5)	38.1 (3.1)	51.9 (2.9)	91
Mixed	10.7 (1.5)	13.4 (2.2)	26.2 (3.2)	80
Proofreading				
Unmixed	3.4 (0.5)	3.4 (0.8)	23.4 (4.9)	53
Mixed	6.2 (0.9)	3.8 (1.0)	13.2 (3.2)	68

^aThe value of p(error on the|error) expected by chance alone is 27.5.

Most critical is the pattern of results for the conditional percentages, which, unlike the absolute percentages of errors, should not be affected by any speed-accuracy tradeoff. The conditional percentages were significantly larger for detection (mean = 39%) than for proofreading (mean = 18%), $F(1,39) = 66.0$, $MS_e = 630$, $p < .001$, and significantly larger for the unmixed passage (mean = 38%) than for the mixed passage (mean = 20%), $F(1,39) = 59.5$, $MS_e = 520$, $p < .001$. In addition there was a significant interaction between these two factors, $F(1,39) = 11.3$, $MS_e = 520$, $p = .002$, reflecting the larger difference between the unmixed and mixed passages in detection than in proofreading (although the difference was in the same direction in both tasks). The only passage for which the conditional error percentage was significantly greater than the chance level (27.5%) was the unmixed passage employed in the t-detection task, $t(90) = 8.2$, $p < .001$.

It is interesting to note in Table 3 that, for the mixed z passage, the conditional error percentage was actually significantly less than chance, $t(67) = 4.8$, $p < .001$. In other words, subjects were less likely to make an error on zhe than on other letter strings containing a z. One possible explanation for this finding is based on the fact that for the mixed z passage there are two factors--mixed typecase and misspellings--each of which should disturb the use of reading units larger than letters. It is reasonable to expect that the combined effect of the two factors would be larger than the effect of each factor individually. For a pure letter scanning (z-scanning) strategy, one might expect a conditional percentage less than chance on the

basis of the finding by Healy (1976, Experiment 1) that the conditional percentage of t-detection errors in the locations was significantly less than chance in a passage of scrambled letters⁵. This finding can be understood by noting that the t in the word the occurs in the first position of the word, and Corcoran (1966) and Smith and Groat (Note 1) observed that the position of the letter within a word affects letter detection and that early letters are more likely to be detected than later ones.

In any case, the major results are in line with the predictions that subjects would automatically process the word the in units larger than the letter in the unmixed passage but not in the mixed passage, and that subjects could not automatically process the letter string zhe as a single unit, as suggested by the unitization hypotheses. In contrast, these results are not consistent with the redundancy hypothesis.

EXPERIMENT 3

According to the unitization hypotheses, the pattern of errors on the in letter detection and substitution proofreading can be attributed to the fact that the is an extremely common word. On the basis of these hypotheses, then, one should see an analogous pattern of results when comparing common and rare words in letter-detection and substitution-proofreading tasks as seen when comparing the and other words. Specifically, subjects should make more errors on common than rare words in letter detection but not in substitution proofreading. We tested this prediction in Experiment 3 by comparing common and rare nouns. Word frequency was controlled in this experiment across the lengths of the words and the locations of the target letter. Each word occurred only once in each of the passages of Experiment 3, so that any differences between common and rare words could not be attributed to different numbers of occurrences of these words in the test passage, a factor that was not controlled in Experiments 1 and 2. Further, in Experiment 3, unlike Experiments 1 and 2, a passage of scrambled nouns was used, so that syntactic and semantic redundancy was virtually eliminated.

Experiment 3 allowed us to test another hypothesis as well, namely the standard explanation of proofreaders' errors: "the common belief that misspellings are more difficult to detect in more familiar words, presumably due to incomplete processing of, or inattention to, orthographic features" (Krueger & Weiss, 1976, p. 204). In a letter-search task with mutilated targets, not unlike the present substitution-proofreading task, Krueger and Weiss (1976) provided support for this hypothesis. In particular, they found that subjects were more likely to detect a mutilated target (created by changing E to F) when it occurred in a nonword than when it occurred in a word. On the basis of these results, we would expect subjects to make more substitution-proofreading errors on common than on rare words in Experiment 3.

Method

Subjects. The same subjects were employed as in Experiments 1 and 2. The subjects performed Experiment 3 concurrently with Experiment 2 (see below).

Design and materials. The two passages used in Experiment 3 had the same punctuation as those used in Experiment 2; only the words differed. The words employed in the first passage ("i passage") were taken from a list of nouns composed by Paivio, Yuille, and Madigan (1968). The passage included 50 common nouns (AA on the Thorndike-Lorge scale, 1944) and 50 rare nouns (5 or less on the Thorndike-Lorge scale). There were 40 *is* in this passage; 20 in common words and 20 in rare words. The words were selected with the following constraint: For every common word, a rare word was chosen that was the same length. Wherever a *i*, if any, occurred in the common word, a *i* occurred in the same location in the corresponding rare word. For example, the common word *fact* was matched with the rare word *pa*i*ct*. There were two versions of this passage (versions A and B), which included the same words but differed in the order of the words. Wherever a common word containing a *i* occurred in one version, its rare mate occurred in the other version. Thus, for example, *fact* in one version was replaced by *pa*i*ct* in the other version, and *pa*i*ct* was replaced by *fact*. The order of the words was otherwise random and the same for both versions. The two versions of this passage were identical to those used by Healy (1976, Experiment 4).

The second passage ("z passage") was the same as the *i* passage except that every *i* was replaced by a *z*. There were no other *zs*. No letter strings containing *z* in the *z* passage formed English words except one which formed a very infrequent word (*ra*z*ing* from *ra*i*ng*)⁶. There were two versions of the *z* passage, which corresponded to the two versions of the *i* passage.

Each subject was given mimeographed copies of both passages, typed on separate sheets of paper. The *i* passage was placed between the two *i* passages for Experiment 2, and the *z* passage was placed between the two *z* passages for Experiment 2. A given subject was shown only one version (A or B) of the passages. The version of the *i* passage shown to a given subject matched the version of the *z* passage shown to him or her. This division among subjects was orthogonal to the three divisions of the subjects made for Experiment 2, so that there were approximately equal numbers of subjects (5-7) in the sixteen (2X2X2X2) subgroups of subjects.

Procedure. The procedure was the same as that used in Experiment 2.

Results and Discussion

The results are summarized in Table 4, which includes for both passages the mean and standard error of the mean percentages of errors on the common and rare nouns. More errors were made on common nouns than on rare nouns for *i* detection, but a small difference in the opposite direction was found for proofreading, for which, overall, errors were less frequent. An analysis of variance conducted on these data revealed that the main effect of task, $F(1,95) = 56.0$, $MS_e = 73$, $p < .001$, the main effect of word frequency, $F(1,95) = 5.2$, $MS_e = 30$, $p = .023$, and the interaction of these two factors, $F(1,95) = 7.0$, $MS_e = 37$, $p = .009$, were all significant.

Table 4

Means and Standard Errors of Means for Letter-Detection (l-circling)
and Substitution-Proofreading (g-circling) Error Percentages
in Experiment 3 as a Function of Word Frequency

Task	Error percentage	
	M	SEM
Detection		
Common	14.2	1.2
Rare	11.3	1.1
Proofreading		
Common	6.2	0.8
Rare	6.4	0.6

This pattern of results is consistent with the unitization hypotheses but cannot be explained by the redundancy hypothesis. Finding fewer errors in proofreading on common than on rare words is also inconsistent with the proposal by Krueger and Weiss (1976) that misspellings are more difficult to detect in more familiar words and with the demonstration by them that mutilated targets were more often missed in a letter-search task when they occurred in words than in nonwords. There were many procedural differences between the present substitution-proofreading task and the letter-search task of Krueger and Weiss (1976). Perhaps the most important difference between the two studies is that the mutilation of the target in the study by Krueger and Weiss (1976) (changing an E to an F) was much smaller than the mutilation of the target in the present study (changing a l to a g). In fact, Krueger and Weiss (1976) proposed that the level of target mutilation may determine whether the mutilated target will be assimilated into the familiar word schema, becoming more difficult to detect, or will be contrasted with the familiar word schema, becoming easier to detect.

SUMMARY AND CONCLUSIONS

In summary, the pattern of errors in this study for proofreading was quite different from the pattern for letter detection. Whereas subjects made an inordinate number of errors on the in letter detection, the number of errors on the was no greater than chance in proofreading, and, in fact, was significantly less than chance in Experiment 1. Likewise, whereas subjects made more errors on common than on rare words in letter detection, a small difference in the opposite direction was found in proofreading. These results provide clear evidence that subjects do not skip over or give inadequate attention to the word the in reading prose, thereby refuting the redundancy

hypothesis. In contrast, these results are consistent with the unitization hypotheses put forth by Drewnowski and Healy (1977). In particular, they support the notion that in reading normal prose subjects are able to process automatically common words, especially the most common word the, in units larger than the letter. When the formation of these larger units is disturbed, as it is when every other letter is typed in capitals or when misspellings are introduced, the subjects are more likely to complete the processing of the words at the letter level and, hence, are less likely to make letter-detection errors on the words.

Another, less attractive, explanation is available for the difference between the pattern of results in the comparable substitution-proofreading (z-circling) and letter-detection (t-circling) tasks of Experiments 2 and 3: Whereas subjects may have read the passages for meaning when performing the letter-detection task, subjects may have been able to scan the text for letters (zs), ignoring meaning altogether, when proofreading. (Although a pure letter scan may not be a reasonable strategy in many proofreading situations, it would be reasonable in the particular substitution-proofreading task used in Experiments 2 and 3, since subjects knew that all misspellings involved z.) In the case of a pure letter scan, one would expect to find no differences between common and rare words, as indeed was the case for proofreading. However, three factors argue against such an explanation. First, such a letter-scanning strategy is impossible in the transposition-proofreading task of Experiment 1, which yielded results consistent with those of Experiments 2 and 3. In the task of Experiment 1, in which misspellings consisted of transpositions, the subjects were forced to access the lexicon in order to determine whether a given letter string included a misspelling. Second, the same subjects performed the letter-detection and substitution-proofreading tasks, and the subjects did the two tasks in immediate succession. Since the tasks were superficially strictly analogous, it would seem unlikely that the subjects would employ radically different strategies in the two tasks. In support of this argument, the order in which the two tasks were performed was not found to be a factor that influenced error frequency.⁷ The third argument against this explanation is that the major portions of the text were identical in the passages for the two tasks. Sixty of the 100 words in the substitution-proofreading task did not contain misspellings so were identical in all respects to the analogous words in the detection task. It seems unreasonable that subjects would process these words in a different way in the two tasks.

Even if an explanation in terms of strategy differences in the two tasks could not be ruled out, the difference between the pattern of results for comparable substitution-proofreading and letter-detection tasks would be of interest. One would still be left with the interesting question: Why were the subjects able to use the more efficient (in terms of numbers of errors) letter-scanning strategy in the substitution-proofreading task but not in the comparable letter-detection task? The only difference between the two tasks (apart from the trivial difference between the identities of the target letters--t versus z) was that the target letters occurred within real words in letter detection but not in substitution proofreading. The question would then be: Why were the subjects able to use the more efficient letter-scanning strategy when the target letters did not occur within real words but not when the target letters did occur within real words? The most plausible answer to

such a question again seems to be in terms of the size of the reading unit available to the subjects. When and only when units are available at the word level may subjects fail to use a pure letter-scanning strategy. Hence, even under the assumption of a strategy difference between tasks, there is support for the central unitization hypothesis.

A possible explanation for the fact that subjects did not make many errors on the in the proofreading tasks of Experiments 1 and 2 is that a large percentage (27.5%) of the misspellings occurred on the in the passages used in these experiments. It could be argued that because of the preponderance of misspellings involving the, subjects gave more attention to that word than they would have otherwise. However, such an explanation could not account for the fact that subjects made a disproportionately large number of errors on the in the t-detection task of Experiment 2, even though an equally large percentage (27.5%) of the ts occurred in the in the passages employed for that task. In addition, this explanation cannot account for the pattern of proofreading errors in Experiment 3, since each word occurred only once in the passages for that experiment. Although a number of ad hoc explanations, like this one, could be constructed to account for a subset of the present results, it appears that only the unitization hypotheses are able to account for the full range of results presented here.

Although this study gives further support to the notion that subjects may read common words in units larger than the letter, the nature of these reading units has not been further clarified by this study. As Healy (1976) has remarked, the units may be perceptual (visual) units or response (phonetic) units. On one hand, the possibility that response units, presumably formed by phonetic recoding, are at issue, rather than visual units, is supported by Corcoran's (1966) study of e detection and subsequent follow-up studies by Mohan (1978), Chen (1976), and Locke (1978), which demonstrated more letter-detection errors on silent than on pronounced letters for normal adult subjects, suggesting that these subjects scan a phonetic representation when searching for a target letter. (However, in a letter-detection experiment by Smith and Groat, Note 1, and in an unpublished experiment by Venezky reported in Hatch, Polin, and Part, 1974, the effect of silent versus pronounced letters was not replicated.) On the other hand, the effects of typecase demonstrated in Experiment 2 of the present study and in Experiment 3 of the study by Drewnowski and Healy (1977) suggest that visual units may be at issue.

Finally, it may be argued that these letter-detection and proofreading tasks have little to do with normal reading for meaning. However, a recent developmental study by Drewnowski (1978) demonstrated that the tendency to make letter-detection errors on the word the was a function of reading level; the pattern of errors on a series of letter-detection tasks provided a good index of the subject's reading ability, presumably because it provided a good index of the size of the reading units employed by the subject. Mohan (1978) has also demonstrated an increasing tendency to make letter-detection errors on the with increases in grade level. These studies demonstrate the relevance of letter-detection tasks to normal reading. If we can understand why subjects make a preponderance of letter-detection errors on the word the, we may indeed advance our understanding of the reading process.

REFERENCE NOTE

1. Smith, P. T., & Groat, A. Spelling patterns, letter cancellation and the processing of text. Paper presented at Processing of Visible Language, Eindhoven, Netherlands, September 5, 1977.

REFERENCES

- Chen, K. Acoustic image in visual detection for deaf and hearing college students. Journal of General Psychology, 1976, 94, 243-246.
- Corcoran, D. W. J. An acoustic factor in letter cancellation. Nature, 1966, 210, 658.
- Corcoran, D. W. J. Acoustic factor in proof reading. Nature, 1967, 214, 851-852.
- Drewnowski, A. Detection errors on the word the: Evidence for the acquisition of reading levels. Memory & Cognition, 1978, 6, 403-409.
- Drewnowski, A., & Healy, A. F. Detection errors on the and and: Evidence for reading units larger than the word. Memory & Cognition, 1977, 5, 636-647.
- Hatch, E., Polin, P., & Part, S. Acoustic scanning and syntactic processing: Three reading experiments--first and second language learners. Journal of Reading Behavior, 1974, 6, 275-285.
- Healy, A. F. Detection errors on the word the: Evidence for reading units larger than letters. Journal of Experimental Psychology: Human Perception and Performance, 1976, 2, 235-242.
- Holbrook, M. B. Effect of subjective interletter similarity, perceived word similarity, and contextual variables on the recognition of letter substitutions in a proofreading task. Perceptual and Motor Skills, 1978, 47, 251-258. (a)
- Holbrook, M. B. Effect of subjective verbal uncertainty on perception of typographical errors in a proofreading task. Perceptual and Motor Skills, 1978, 47, 243-250. (b)
- Krueger, L. E., & Weiss, M. Letter search through words and nonwords: The effect of fixed, absent, or mutilated targets. Memory & Cognition, 1976, 4, 200-206.
- Kučera, H., & Francis, W. N. Computational analysis of present-day American English. Providence, R. I.: Brown University Press, 1967.
- LaBerge, D., & Samuels, S. J. Toward a theory of automatic information processing in reading. Cognitive Psychology, 1974, 6, 293-323.
- Locke, J. L. Phonemic effects in the silent reading of hearing and deaf children. Cognition, 1978, 6, 175-187.
- Mohan, P. J. Acoustic factors in letter cancellation: Developmental considerations. Developmental Psychology, 1978, 14, 117-118.
- Paivio, A., Yuille, J. C., & Madigan, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. Journal of Experimental Psychology Monograph, 1968, 76 (1, Pt. 2).
- Rayner, K. Visual attention in reading: Eye movements reflect cognitive processes. Memory & Cognition, 1977, 5, 443-448.
- Schindler, R. M. The effect of prose context on visual search for letters. Memory & Cognition, 1978, 6, 124-130.
- Shiffrin, R. M., & Schneider, W. Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. Psychological Review, 1977, 84, 127-190.
- Smith, F. Understanding reading: A psycholinguistic analysis of reading and learning to read. New York: Holt, Rinehart, and Winston, 1971.
- Taylor, W. L. "Cloze procedure": A new tool for measuring readability.

Journalism Quarterly, 1953, 30, 415-433.
Thorndike, E. L., & Lorge, I. The teacher's word book of 30,000 words. New York: Columbia University, Teachers College, Bureau of Publications, 1944.

FOOTNOTES

¹The fact that words were created by these transpositions seems inconsequential, since none of the 96 subjects made an error on the word fro and only three made errors on the word eh.

²Word length was confounded with two potentially important variables: word frequency and the location of the transposed letters in the word. The mean frequency (Kuřera & Francis, 1967) of the misspelled words excluding the monotonically decreased as a function of word length. Mean frequency per 1,014,232 words of text was 18,972, 3,820, 1,625, 70, and 69 for the misspelled words of length two, three, four, five, and six to ten letters, respectively. However, word frequency cannot account for finding a greater mean percentage of errors on the most frequent word the (frequency = 69,971) than on other misspelled three-letter words or finding a greater mean percentage of errors on two-letter words than on three-letter words ($p < .01$ by a Newman-Keuls test).

The location of the transposed letters in a word may be critical, since the transpositions necessarily involved an end letter (either initial or terminal) in all the two- and three-letter words but involved only intermediate letters in three of the six four-letter words, two of the four five-letter words, and seven of the nine words six to ten letters long. For the four- and five-letter words, more errors were made on the five words with transpositions involving only intermediate letters (mean = 21.5%) than on the five words with transpositions of an end letter (mean = 4.6%), $F(1,95) = 97.4$, $MS_e = 140$, $p < .001$. However, other factors must also be critical, since the percentage of errors on the long word separate when misspelled as esparate was quite high (21.9%), although it involved a transposition of the initial letter.

³The difference between errors percentages on the word the and other three-letter words is due in part to the relatively large percentage of errors (14.6%) made on a single instance of the word the (the only instance involving a capital letter: The was misspelled as Hte). Excluding that instance, the mean percentage of errors on the was greatly reduced (mean = 2.2%). The especially low percentage of errors on other three-letter words (only two errors out of 384 opportunities across subjects) may be due to peculiar aspects of the particular misspellings employed: for was misspelled as fro, let as lte, low as olw, and two as wto. However, with the exclusion of the one instance of the described above, the difference between errors made on the misspelled as hte (mean = 2.3%) and teh (mean = 2.1%) was not significant, $F(1,95) < 1$.

⁴In the analysis of variance, there were missing cells for the cases when a subject made no errors in a given task on a given passage so that a conditional percentage could not be computed. Each of these missing cells was replaced by the appropriate mean conditional percentage. A conservative estimate of the degrees of freedom was computed by subtracting the number of

subjects (56) who contributed one or more missing cells.

⁵In Healy's (1976) experiment the scrambled letter passage was derived from a prose passage by retaining the punctuation, word boundaries, and locations of the ts in the prose passage but scrambling all the remaining letters. A t in a the location in the scrambled letter passage was a t in a location where the word the occurred in the corresponding prose passage.

⁶The fact that a word was created in this case seems relatively inconsequential, since only 6 of the 96 subjects made an error on the word razing.

⁷Unweighted analyses of variance for unbalanced designs (unequal cell frequencies) were performed on the total error scores in Experiments 2 and 3. The factor of test order (proofreading first versus letter detection first) was not found to be a significant main effect or to enter into any significant interactions in these analyses (except for a significant, $p = .040$, five-way interaction in Experiment 3, involving the four between-subjects factors and the factor of task.)

THE SPEECH CODE AND LEARNING TO READ*

Donald Shankweiler,⁺ Isabelle Y. Liberman,⁺ Leonard S. Mark,⁺
Carol A. Fowler,⁺⁺ and F. William Fischer⁺⁺⁺

Abstract. Good and poor readers among second-grade school children can be distinguished by the extent to which their recall of random letter strings is affected by the phonetic characteristics (rhyming or not rhyming) of the items. The recall performance of mildly backward (marginal) readers was less penalized by phonetic confusability than that of superior readers, and severely backward (inferior) readers showed a still weaker effect of confusability. These results were obtained not only for visual presentation of the letter strings (Experiments 1 and 2), but also for auditory presentation (Experiment 3). Taken together, the findings support the hypothesis that good and poor readers differ in their use of phonetic coding in working memory, whatever the sensory route of access, and they suggest that individual variation in coding efficiency may be a relevant factor in learning to read. It is suggested that a number of memory-related problems typical of poor readers may be manifestations of deficiencies in phonetic coding.

INTRODUCTION

In the research presented here, we explore the possibility that children who learn to read with facility differ from those who learn to read with difficulty in the extent to which they rely on speech-related processes in short-term memory. We have supposed that a major function of speech coding in reading is its use in comprehension of stretches of text longer than the word. Thus, our concern is directed toward the role of the phonetic representation as a medium for linguistic storage.

It is obvious that perception of language, whether written or spoken, requires that a reader or listener hold a sufficient number of individual words and their order of arrival long enough to permit interpretation of each sentence. There is a wealth of evidence (Baddeley, 1966; Conrad, 1964, 1972; Conrad & Hull, 1964; Hintzman, 1969) that for this purpose, the working memory, in both reading and listening, may rely on phonetic coding of the

*To be published in Journal of Experimental Psychology: Human Learning and Memory.

⁺Also at University of Connecticut, Storrs.

⁺⁺Also at Dartmouth College, Hanover, N. H.

⁺⁺⁺Now at Central Connecticut State College, New Britain.

Acknowledgment: This work was partially supported by a grant from the National Institute of Child Health and Human Development (Grant HD-01994) to Haskins Laboratories.

information to be retained. Whether the information is letters, words or syllables, it is consistently found that confusions in recall are greater when the items are phonetically similar than when the similarity is either visual or semantic. This suggests that perceivers have so strong a tendency to store the information in phonetic form that they persist in using this form of coding even when it penalizes recall. Strikingly parallel results have been obtained when words are presented logographically as Japanese kanji characters (Erickson, Mattingly, & Turvey, 1977) or as Chinese characters (Tzeng, Hung, & Wang, 1977) suggesting that it may benefit a reader to recode phonetically regardless of whether he uses an alphabet or a logographic writing system. Moreover, even when the stimuli are not linguistic items at all, but pictured objects, there is evidence that the information may nevertheless be recoded phonetically in memory (Conrad, 1972). Together, all these findings underscore the general use of phonetic coding as a widely applicable strategy for holding in temporary storage any information that can be linguistically processed.

We must consider, of course, the possibility that some readers may employ different kinds of working memory representations than listeners do. In principle, the possibility certainly exists that a nonphonetic representation of a visual or semantic kind might be used. Indeed there is evidence that nonphonetic strategies are employed by some congenitally deaf readers (Frumkin & Anisfeld, 1977; Locke, 1978). But the well-attested difficulties of congenitally deaf children in learning to read (Swisher, 1976) also suggest that nonphonetic strategies may not work well.

Although it may be inferred that nonphonetic strategies are less common than phonetic ones for the normal adult, there is little information on children at the point of learning to read. We assume that successful beginning readers (of English), who have learned to relate the structure of the printed word to the phonological and phonetic structure of the spoken word,¹ have the phonetic form of the word available for use in working memory. Poor readers, on the other hand, have difficulty in employing this analytic strategy, as the nature of their reading errors shows (Shankweiler & Liberman, 1972; Fowler, Liberman, & Shankweiler, 1977). Consequently, like some of the congenitally deaf, poor readers may tend to rely more on nonphonetic strategies in working memory.

The possibility that differences in children's use of phonetic coding may be related to success or failure in learning to read has only recently begun to be explored² (Liberman, I. Y., Shankweiler, Liberman, A. M., Fowler, & Fischer, 1977; Shankweiler & Liberman, I. Y., 1976). In consideration of the use of phonetic coding in the working memory (Baddeley, 1978; Baddeley & Hitch, 1974; Crowder, 1978; Kleiman, 1975; Levy, 1977), and in recognition of differences in the characteristic strategies of word recognition employed by successful and unsuccessful beginning readers, it seemed worthwhile to ask whether beginning readers who are progressing well can be distinguished from those who are doing poorly by the degree to which they rely on phonetic coding in a task designed to stress working memory.

A task was selected in which the effects of phonetic coding are readily detected. We borrowed a procedure devised by Conrad (1972) for adult subjects in which performance is compared on recall of phonetically similar (rhyming)

and phonetically dissimilar (nonrhyming) sequences of letters. It was expected that the phonetically similar items would generate confusions and thus penalize recall in subjects who use a phonetic code. If poor readers were deficient in the use of a phonetic code, they might be expected to be less affected by the phonetic similarity of the items than good readers, whether or not the groups differ in recall of phonetically dissimilar items.

Experiment 1

Method

Subjects. The subjects were school children who were nearing completion of the second school year at the time the experiment was conducted. Reading teachers were asked to select the best and the poorest readers in their respective classes. These children were then given the word recognition subtest (Jastak, Bijou, & Jastak, 1965) of the Wide Range Achievement Test (WRAT) and a test of intelligence (Dunn, 1965), the Peabody Picture Vocabulary Test (PPVT). On the basis of scores obtained on the WRAT, three groups were selected that were nonoverlapping in reading level. Table 1 gives the particulars for each group.

Table 1

Estimated mean reading grade,* mean age and IQ[†] for second grade children grouped according to reading attainment.

Group	n	age	IQ	Reading Grade
Superior	17	8.0	113.9	4.9
Marginal	16	8.1	101.7	2.5
Poor	13	8.2	111.6	2.0

*Reading grade equivalent score on reading subtest of the Wide Range Achievement Test.

[†]Peabody Picture Vocabulary Test.

As may be seen from the table, the first group, designated as the superior readers, was composed of 17 children who were reading well ahead of grade placement, having obtained a mean grade equivalent of 4.9 on the WRAT. The second group, the marginal readers, included 16 children who averaged slightly less than one-half year of retardation in reading (having obtained a

mean WRAT equivalent of 2.5). The third group, containing 13 children whom we called poor readers, obtained a mean WRAT equivalent of 2.0, indicating nearly a full year of retardation in reading.

The three groups did not differ significantly in mean age. In each, the mean IQ level as assessed by the PPVT was above 100. The means were closely matched for the two extreme groups. The marginal readers averaged about 10 score points below the others, a fact that, in view of the results obtained, could not be of great importance.

Stimuli: Simultaneously-presented letter strings. Sixteen strings of five upper-case letters were devised for presentation by projector tachistoscope. Eight of the five-letter strings were composed of rhyming consonants (drawn from the set B C D G P T V Z) and eight were composed of nonrhyming consonants (drawn from the set H K L Q R S W Y). In generating the test sequences each letter was allowed to appear only once in a given sequence and all letters appeared equally often in each serial position. The rhyming and nonrhyming sets were interleaved and all 16 sequences were randomized. The test sequences were preceded by an identification test in which each of the 16 consonant letters was presented individually, centered on the screen, twice each in randomized order.

A 2 x 2 inch slide was constructed for each of the 16 test sequences. Each typed letter string was centered on the slide, the group of 5 letters subtending a visual angle of 4.8 degrees horizontally when projected on the viewing screen for a viewing distance of 11 feet. The slides were displayed using a slide projector equipped with a projector tachistoscope which was controlled by a bank of three 100 sec timers.

Procedure. The subjects were tested in groups of approximately 15 children. First, the identification pretest was given. Each pretest trial was preceded by an alerting stimulus, an asterisk, centered in the display field and shown for 1 sec. The stimulus followed 1 sec after the asterisk was turned off. Each letter was then displayed for 1 sec, after which the children were allowed as much time as needed to write the letter on the answer sheet. After completion of the pretest, the children were told that they were about to see groups of letters, and that their task was to write the letters in the order given when the experimenter said "write." The procedure for the test trials was the same as in the pretest except that each five-letter stimulus item was displayed for 3 sec.

The test was given twice: once with immediate recall, and once with delayed recall. Three practice trials introduced each condition. In the first condition, the children were requested to write their responses immediately following each exposure. In the delay condition, 15 sec elapsed between tachistoscope presentation and the signal to respond. The children were requested to sit quietly during the delay interval; no intervening task was imposed. Half the subjects began the test session with the immediate recall condition, while the remainder started with the delayed recall. The children recorded their responses on an answer sheet containing rows of five underlined blank spaces.

Table 2

Mean Errors Summed Over Serial Positions (max = 40) for Simultaneous Presentation of Visual Sequences (V1) in Experiment 1 and for Successive Presentation of Visual (V2) and Auditory (A) Sequences in Experiments 2 and 3

		Immediate Recall						Delayed Recall					
		Confusable			Nonconfusable			Confusable			Nonconfusable		
		V1	V2	A	V1	V2	A	V1	V2	A	V1	V2	A
Superior M		20.1	20.9	22.4	10.8	10.0	6.3	29.1	27.7	26.9	10.8	11.6	8.4
Readers	s.d.	(5.8)	(6.4)	(6.2)	(7.0)	(7.9)	(5.7)	(5.1)	(5.6)	(4.4)	(6.5)	(7.7)	(7.0)
Marginal M		25.1	26.8	25.9	19.9	22.2	18.0	31.1	31.3	31.7	25.1	24.4	23.2
Readers	s.d.	(4.4)	(6.8)	(7.0)	(5.2)	(9.0)	(9.7)	(5.1)	(4.5)	(4.0)	(8.5)	(10.7)	(9.2)
Poor M		28.4	31.4	31.7	24.2	24.2	21.4	30.5	35.1	34.7	25.8	24.2	25.2
Readers	s.d.	(1.9)	(4.8)	(3.4)	(6.0)	(8.2)	(7.0)	(4.6)	(3.0)	(2.6)	(5.9)	(8.8)	(6.2)

Figure 1. Mean recall errors summed over serial positions. Means from delay and nondelay conditions are averaged.

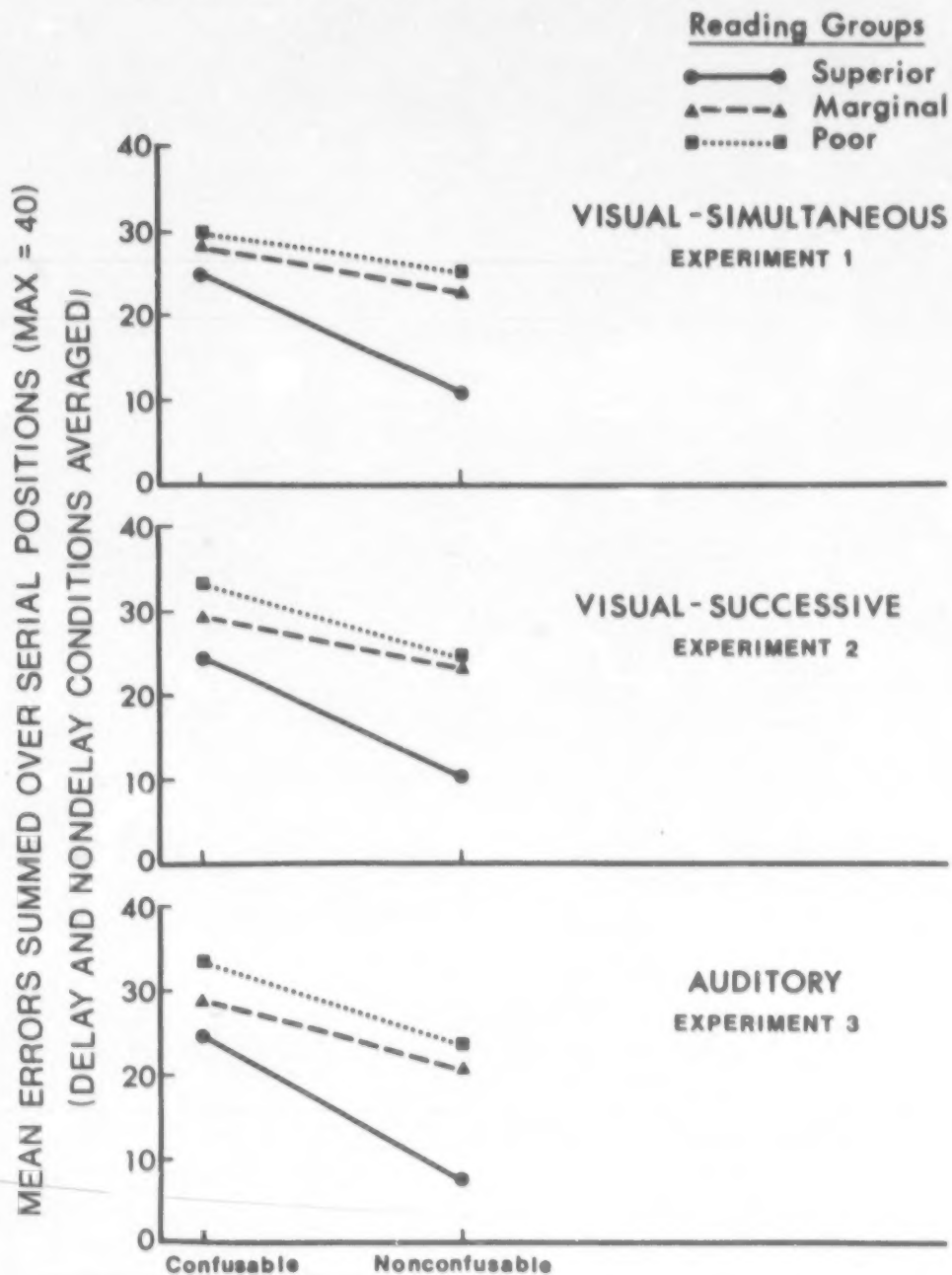
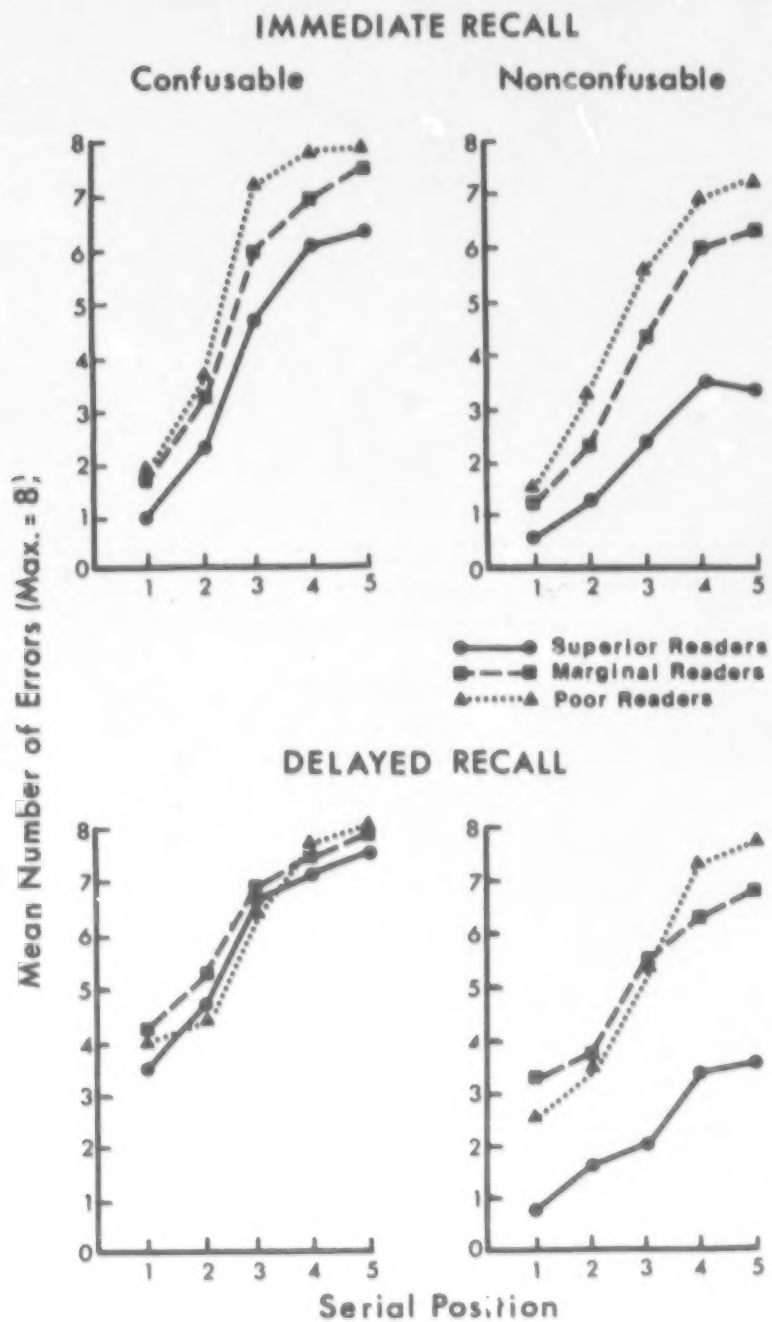


Figure 2. Mean recall errors for the visual-simultaneous condition (Experiment 1) replotted as a function of serial position.



Results and Discussion

Few errors occurred in identification of single letters on the identification pretest, and there were no significant differences among the reading groups. We assumed, therefore, that any differences between the groups in their performance on the experiment proper are not attributable to differential accuracy in letter identification.

Serial recall of the five-item sequences was scored in two ways. The first, more stringent, procedure counted a response as an error if it was an incorrect item identification or if it was a correct identification assigned to the wrong serial position. The second procedure counted a response as an error only if it was not a member of the stimulus string. In either case, the dependent measure for each subject was his total number of errors summed over serial positions.

We first consider the analysis based on the more stringent scoring procedure. The error data for this analysis in Experiment 1 and in the subsequent experiments are summarized in Table 2. In order to show the overall effect of phonetic confusability more clearly, Figure 1 displays the results averaged across the delay and nondelay conditions. It is apparent from inspection of Figure 1 (top graph) that phonetic similarity exerts an effect, but that the effect is much stronger for superior readers than for poor ones, with marginal readers falling in between.

These effects were substantiated by a three-way factorial analysis of variance, in which the between-groups factor is reading achievement, and the within-groups factors include item type (i.e., rhyming or nonrhyming letter names), and delay or nondelay of response. The overall effect of reading group is significant, $F(2, 43) = 22.7$, $p < .001$, in the expected direction: Superior readers made fewer errors than the others. We see at once that the main differences are between the superior readers and the other groups; marginal and poor readers did not differ significantly from each other.

In accord with many past findings on adults, the phonetic characteristics of the items markedly influenced the rate of correct recall of the children as a whole, that is, there was a significant main effect of item type, $F(1, 43) = 73.0$, $p < .001$. But of particular interest from our standpoint is the fact that there were notable differences in the effects of phonetic similarity on the recall of children who differed in reading level. Thus we find a significant interaction between reading group and item type, $F(2, 43) = 9.9$, $p < .001$. It is apparent from the figure that the superior readers are more adversely affected by item confusability than the other groups.

So far we have examined the gross aspects of the response pattern, having considered errors summed over serial positions and averaged over the delay and nondelay conditions. Figure 2 displays errors for each serial position separately for nondelay (top) and delay (bottom) conditions. Two facts stand out on inspection of the figure. First, a strong effect of serial position is present in the data of all three reading groups. Second, Figure 2 makes readily visible a fact already mentioned in discussion of the top graph of Figure 1—namely, that the superior readers were more strongly penalized than the inferior readers by phonetic confusability among the stimulus items.

We may now see that the penal effect of phonetic confusability on the good readers is magnified by delay of recall. Delay leads to an overall increase in errors, $F(1, 43) = 29.8$, $p < .001$, but its effect is marked only in superior readers. The interaction between group and delay is not significant, $F(2, 43) = 2.3$, $p = .113$, but when we take into consideration the additional factor of item type, the three-way interaction is significant, $F(2, 43) = 8.2$, $p < .001$. It is apparent from inspection of the figure that the interaction is due to the departure of the superior group's performance from the performance of the poor and the marginal readers. It may be seen in the lower portion of Figure 2 that the superior readers are sharply distinguished from the others in recall of phonetically nonconfusable items, and nearly indistinguishable in their recall of confusable items.

In view of speculation that a function of the phonetic representation in working memory is to preserve information about serial order, it is of interest to ask to what extent this pattern of results reflects errors of serial order alone, and to what extent it reflects forgetting of the items. The data obtained from the less stringent scoring procedure (in which responses were scored without regard for serial order) were subjected to an analysis of variance parallel to that described above. The results can be stated briefly. All the main effects and interactions that were significant in the analysis of the data in which serial order was taken into account were also significant when the order in which the subject wrote down the responses was ignored. We interpret this to mean that the pattern of results reflects forgetting of items, not merely errors of serial order.

To summarize the findings of Experiment 1, we have seen that superior readers were clearly better at recall of phonetically nonconfusable items than were the poor readers, while at the same time failing to show a clear advantage on the confusable items. We regard this as an interesting result. It is a relatively easy matter to demonstrate that poor readers do less well than good readers on a variety of language-dependent tasks. But here, by manipulation of the phonetic characteristics of the test items, we have virtually eliminated the advantage of the superior readers.

It might be supposed, following a line of thought adopted by Bakker (1972) and Corkin (1975), that poor readers suffer specifically from a difficulty in reproducing the order of the items in the memory set. Although this idea may have some merit, our results suggest that the difference between our groups of good and poor readers cannot be attributed solely to differences in susceptibility to order confusions, since the pattern of the results was much the same when the scoring credited the correctly recalled items regardless of whether or not they were recalled in the correct serial position.

The results of Experiment 1 bear out our expectation in demonstrating significant differences in susceptibility to phonetic confusions in working memory among children who differed in reading ability. It may be, as the strongest form of our hypothesis would suppose, that poor readers attempt to hold the items in memory in some nonphonetic form. If they were attempting to use a nonphonetic strategy, they cannot wholly have succeeded since they did show some effect of confusability. Speculation is, in any case, premature until we know the result of presenting the sequences to be recalled to the ear instead of to the eye.

Experiments 2 and 3:
An Auditory Analog and Its Visual Counterpart

From the results of Experiment 1, it could be argued that the problem of the poor readers lies in recoding visual symbolic material into phonetic form. If that is the case, then phonetic confusability of auditorily presented items should affect them neither more nor less than the superior readers. Moreover, even if there were quantitative differences in memory capacity between the two groups, we might still expect that the interaction between reading level and item type (demonstrated in the foregoing experiment) would disappear. If, on the other hand, the interaction remained, then it would follow that the difference between good and poor readers in their use of a phonetic representation is not specifically linked to the visual information channel.

In Experiment 1, it will be recalled, the five-letter sequences were presented in a single exposure. Since auditory presentation requires temporally successive presentation of the items comprising each trial, a parallel visual experiment was required with successive exposure of the items in each letter sequence. The new experiments--the auditory recall task (Experiment 3) and its visual counterpart (Experiment 2)--were made to be as nearly identical as possible except modality of input. We were fortunate in being able to carry out the new experiments on the same subjects who participated in Experiment 1.

Method

Subjects. The children who served as subjects in Experiment 1 were used in Experiments 2 and 3, which were carried out 4-5 months after the original investigation. Two of the poor readers from the original sample had moved away from the area, leaving 11 poor readers. The loss of these subjects did not significantly alter the mean chronological age, IQ or WRAT reading grade of the poor readers (CA = 8.3; IQ = 111.6; WRAT grade equivalent = 2.0).

Visual Successive Task (Experiment 2)

The sequences of letters used in this experiment were the same as those of Experiment 1. However, in the present experiment the letters in a trial were presented successively rather than simultaneously as in Experiment 1. One letter was centered on each slide; thus, five slides were required to form the entire sequence. An additional slide containing an asterisk was inserted at the beginning of the letter slides as a preparatory signal. An identification pretest employed the same slides as were used in Experiment 1.

Procedure. The subjects were tested in groups of no more than six. The instructions and test procedure for the visual identification test were identical to those given in Experiment 1 with the exception of the exposure duration which was, in this case, 500 msec per letter, with an interstimulus interval of 1 sec. Following the identification test, directions for the letter sequences were given. The children were told that on each trial an asterisk would be displayed to signal that a letter sequence was about to appear. At the same time the experimenter operating the tachistoscope would say "ready." Five letters would then be displayed one by one. The children were instructed to write down the letters in the order in which they were

presented. They were instructed to begin writing at the sound of a "clicker," which had been demonstrated previously. In the immediate recall condition the clicker was sounded just as the last letter disappeared from the screen. On the delayed recall task the experimenter waited for a timed interval of 15 sec before sounding the clicker. Three practice trials were given before each recall condition.

The children wrote their responses in booklets, on each page of which was a single line of five dashes corresponding to the five items in a sequence. A separate page in the booklet was used for each sequence. Page colors were alternated so that it could easily be determined that the children were all writing their responses on the appropriate sheet.

Auditory Task (Experiment 3)

The auditory version of the serial recall task was presented to the children on a different day. In most cases a week or more elapsed between the two test sessions. The order of the visual and auditory presentations was counterbalanced.

Stimuli. The stimuli consisted of recorded utterances of the names of the same set of 16 letters that were employed in the two preceding visual experiments. One token of each was recorded on magnetic tape by a male speaker. The recorded utterances were subsequently digitized and edited using the pulse-code modulation system at Haskins Laboratories (Cooper & Mattingly, 1968). The purpose of editing was to equate the duration of the tokens and to adjust the peak amplitudes, making them as nearly equal as possible. The items for an identification pretest were prepared in the same manner.

The stimulus sequences were also constructed with the aid of the PCM system and a timing program designed to output timed sequences of stimuli (Cooper & Mattingly, 1968). A recorded utterance of "ready" preceded the first item of each sequence. The first stimulus token followed 1 sec after the offset of the preparatory stimulus. The interstimulus interval within each sequence of five tokens was also 1 sec. A sequence was terminated by a brief 1000-Hz tone which sounded 250 msec after the offset of the final item in the immediate recall condition, and 15 sec after in the delayed recall condition. The tone served to signal the subjects to begin writing down the preceding sequence. An intertrial interval of 15 sec was programmed to allow ample time for the subjects to record their responses. In the rare instances in which a child required more time, the experimenter stopped the tape between trials. The silent period was broken by the signal "ready" which marked the beginning of the next sequence.

Procedure. The instructions and procedure for the auditory task (Experiment 3) were identical to those employed in the visual sequential task (Experiment 2) with the exception that a tone programmed on the test tape was used to initiate the written responses instead of the sound of a clicker controlled by the experimenter. The children were tested in groups of six or less. Two experimenters were present at each session. One was responsible for reading the instructions to the children and monitoring their behavior during the test; the other operated the tachistoscope or magnetic tape playback. As in the visual experiments, the auditory test sequences were

preceded by an identification pretest.

Results and Discussion

Data from Experiments 2 and 3 were analyzed in a fashion parallel to Experiment 1. Two three-way factorial analyses of variance were performed, one on each set of scores, to evaluate the effects of reading group, item type, immediate vs. delayed recall and the interactions among these variables. A third analysis of variance was carried out to permit direct comparison of the visual and auditory modes of presentation upon recall performance. This was a four-way factorial analysis in which modality, item type, immediate vs. delayed recall and reading group were the variables.

The data of Experiments 2 and 3 are summarized in Table 2 in the columns headed V_2 (visual-successive condition) and A (auditory condition). Each cell in this table gives the mean error score, with its standard deviation averaged across subjects within each group and summed over serial positions. The table permits us to compare the results of the two visual conditions and the auditory condition side by side. These results are remarkable for their similarity across conditions. The visual-successive condition (of Experiment 2) yielded a very similar pattern of results to those of the visual-simultaneous condition of Experiment 1. This was expected. What was unexpected is that auditory presentation resulted in many of the same differences between the performances of good and poor readers as were obtained in the visual conditions.

Visually-presented Sequences (Experiment 2)

Because this was essentially a control experiment for the auditory-successive condition (Experiment 3), we can be brief in our description of the results. They are of interest chiefly in that they replicate so completely the results of Experiment 1, which differed from the present experiment in only one major methodological particular: the group of items to be recalled was presented in a simultaneous display instead of successively one by one.

As in the earlier experiment, each of the main effects of the analysis of variance was significant with $p < .001$. They were as follows: reading groups, $F(2, 41) = 11.9$; item type, $F(1, 41) = 115.3$; immediate vs. delayed recall, $F(1, 41) = 16.4$. We now examine the interactions of interest. Superior readers, as in Experiment 1, are more affected by the phonetic characteristics of the items than the other groups. This is manifested by a significant interaction between reading group and item confusability, $F(2, 41) = 6.5$, $p < .005$. A comparison of the top panel of Figure 1 (the comparable interaction under simultaneous presentation) with the middle panel shows that each interaction effect occurred because the superior readers made fewer errors than the inferior groups on the phonetically dissimilar sequences, whereas the three groups were more nearly at the same performance level on the rhyming sequences.

The analysis does reveal one difference between the two experiments. A feature of Experiment 1 was a significant three-way interaction between reading group, item type and immediate vs. delayed recall, reflecting the fact that delay magnified the differences between the groups, but only on nonrhym-

ing items. This interaction was not obtained in the present experiment, $F(2, 41) < 1$.

Auditorily-presented Sequences (Experiment 3)

It is apparent from inspection of Table 2 that the results of the auditory condition closely paralleled those of the two visual conditions. As in the visual conditions, the factor of the phonetic similarity of the items is a potent one. Each main effect of the analysis of variance was significant at $p < .001$. They are as follows: reading group, $F(2, 41) = 18.7$, item type, $F(1, 41) = 192.2$, immediate vs. delayed recall, $F(1, 41) = 39.2$.

Whether, as with visually presented stimuli, the phonetic characteristics of the items to be recalled affect good and poor readers differently is the major focus in this experiment.³ The analysis shows that this is indeed the case, as revealed by a significant interaction between reading group and item type, $F(2, 41) = 10.7$, $p < .001$. A comparison of the graph of this interaction effect (Figure 1, bottom panel) with the comparable ones from the visual conditions (top and middle panels) shows that the interaction is significant for the same reason as before; the superior readers were more affected by the confusable sequences than were the inferior reading groups.

As in Experiment 2, but not in Experiment 1, there was no significant three-way interaction between reading group, item type and immediate vs. delayed recall, $F(2, 41) < 1$.

We will be aided in making a detailed comparison between Experiment 2 and Experiment 3 by examination of Figure 3. This figure (which is directly comparable to Figure 2) gives mean recall errors for each serial position on each experimental condition. Comparing the graphs in the first column of the figure with those in the second, we see that although the marginal and poor readers did show a degree of phonetic interference, it is clearly of lesser magnitude than that displayed by the superior readers. If we compare the plots in these columns of the figure with the graphs in columns 3 and 4, we see that the pattern is remarkably similar to that obtained in the visual counterpart to this experiment. This point is demonstrated statistically by the analysis of variance in which the factor of immediate vs. delayed recall was collapsed giving a four-factor design in which the factors were modality (visual vs. auditory), reading group, serial position, and item type. In this analysis the modality-by-reading-group-by-item-type interaction failed to approach significance, $F(2, 41) = 1.4$. Thus the factor of phonetic similarity was no less potent in its effect on auditory presentation than on visual.

As in Experiment 1, a set of parallel analyses of variance was carried out on scores derived from an alternative method of scoring in which serial order was disregarded in tallying the items correctly recalled. The outcome is basically the same as that which was reported for Experiment 1. The significant main effects and the interactions that we have considered above all yielded significant effects, in both the auditory and visual conditions.

The principal thing we learn from Experiments 2 and 3, as plainly revealed in Figure 3, is that phonetic similarity produces a differential effect on the recall by good and poor readers whether the items are presented

auditorily or visually. This leads us to a different interpretation of the phenomenon than the one we favored when we had done Experiment 1 but before we had completed Experiments 2 and 3. Our original supposition was that the poor readers' difficulty had to do with recoding from alphabetic characters to a phonetic representation of the linguistic message. Experiments 2 and 3, on the other hand, tell us that poor readers have difficulty accessing or using a phonetic representation whether its origin is print or speech. Hence the problem could not be limited to recoding.

A difference may be noted between the outcome of Experiment 1 and Experiments 2 and 3 in the effects of immediate and delayed recall on the error score. In Experiment 1, we noted that delay magnifies the differences between the reading groups in susceptibility to phonetic confusion in recall. This is manifested in a significant triple interaction in the analysis of variance. However, no interaction was present under either the visual or the auditory conditions of Experiments 2 and 3, respectively. Thus we may be sure that this discrepancy has nothing to do with modality of input. As we shall see, it can plausibly be attributed, instead, to the manner in which the stimuli were presented--i.e., successively or simultaneously, since Experiments 2 and 3 share the characteristic of successive presentation, both in contrast to Experiment 1 in which the group of letters was presented simultaneously.

GENERAL DISCUSSION AND CONCLUSIONS

Possible Consequences of a Deficiency in Phonetic Coding

Each of the three experiments answers yes to the question that motivated our investigation: Can good and poor readers be distinguished by the extent to which their performance on a serial recall task is affected by the phonetic characteristics of the items? Whereas superior readers made considerably fewer errors than poorer readers on the nonrhyming letter strings, the groups were less distinguishable on the rhyming strings. The recall performance of both the mildly backward ("marginal") readers and the severely backward ("inferior") readers was less penalized by phonetic confusability than that of the superior readers in simultaneous visual presentation of the letter strings (Experiment 1), in successive visual presentation (Experiment 2), and in auditory presentation (Experiment 3).

The findings of the three experiments, taken together, support the hypothesis that good and poor readers differ in their use of speech coding, whatever the route of access, and they suggest that individual variation in coding efficiency places limits on reading acquisition. Since differential effects of phonetic confusability on good and poor readers occurred regardless of whether input was to the eye or to the ear, we suspect that difficulties of poor readers are not limited to the act of recoding from script but are of a more general nature. A benefit of this hypothesis is that it permits us to bring together a number of previously unrelated findings regarding the cognitive characteristics of poor readers, and permits us to view the findings as related manifestations of a unitary underlying deficit. It remains for us to examine the expected consequences of a general phonetic coding deficit both within the confines of our experimental task and in the reading process

VISUAL-SUCCESSIVE

EXPERIMENT 2

AUDITORY

EXPERIMENT 3

Immediate Recall

●— Superior Readers
 ▲— Marginal Readers
 ■— Poor Readers

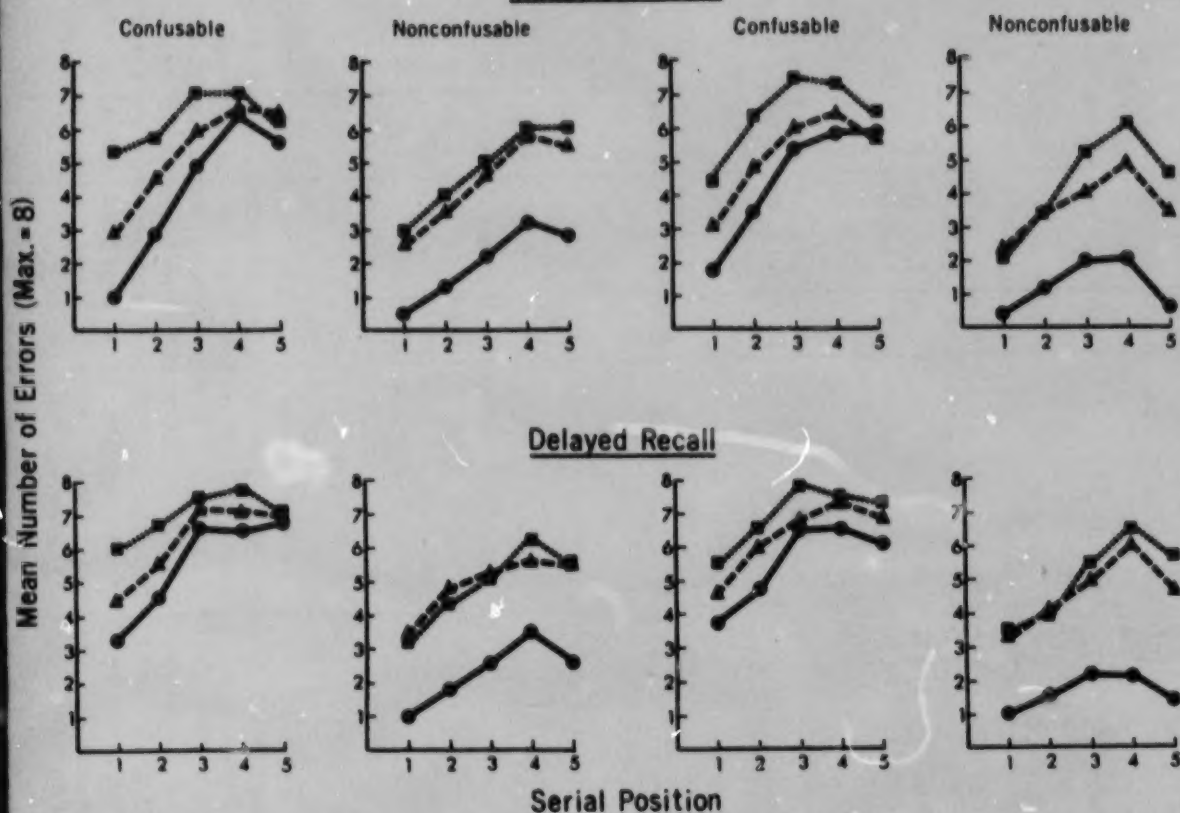


Figure 3. Mean recall errors for the visual-successive condition (Experiment 2) and the auditory condition (Experiment 3) replotted as a function of serial position.

generally.

Rehearsal. A possible manifestation of a general deficiency in the poor readers' use of a phonetic code is slow, ineffective rehearsal of phonetically coded items. Given that the three experiments demanded retention and recall of arbitrary strings of items, it may fairly be said that the situation encouraged rehearsal. Phonetically confusable items could reasonably be expected to generate more interference for good readers than for poor readers if the good readers rehearse confusable items at a more rapid rate.

An interpretation that emphasizes the relevance of rehearsal is compatible with the results of Experiment 1 (visual simultaneous stimulus presentation) in which it was found that imposing a 15-sec delay between stimulus presentation and recall adversely affected the good readers' performance on rhyming items and the poor readers' performance hardly at all. However, these differential effects of delay of recall were absent in Experiments 2 and 3. In these experiments, both groups were adversely affected by delay. The relevant difference is presumably the successive presentation of items, which permits opportunity for rehearsal during delivery of the string. It is possible that successive presentation may tend to evoke an active rehearsal strategy in both good and poor readers.

In any case, an experiment of Mark, Shankweiler, I. Y. Liberman, and Fowler (1977) leads us to believe that differences in efficiency of rehearsal cannot alone account for the differences we obtained between good and poor readers. The subjects in the study of Mark et al., selected in the same manner as those in the present experiments, were tested on rhyming and nonrhyming words using a recognition memory task that minimized the opportunity for rehearsal. As in the present study, the good readers were adversely affected by phonetic similarity among the items to a much greater extent than the poor readers, though rehearsal strategy or its effectiveness could not plausibly have been a distinguishing factor.

However, none of the findings we have mentioned is incompatible with the possibility that a short-term rehearsal loop plays an important part in reading and in learning to read, as Baddeley (1978) has suggested. It would be of interest to discover whether good readers are more susceptible to suppression of rehearsal than poor readers. In this connection, Bauer (1977) has studied matched groups of children with and without "learning disabilities" on recall of word strings in which a delay interval was either unfilled or occupied by a task designed to block rehearsal. Differences between the groups were greater with the unfilled interval, suggesting that the subjects with learning problems did rehearse, but not as effectively as the controls. At any rate, with respect to the proposal that poor readers have a rehearsal problem, we would wish to maintain that underlying the slower (or otherwise less effective) rehearsal of the poor readers may be their poorer access to a phonetic code or their access to a degraded phonetic representation. Thus, from our perspective, the primary problem is the availability of a phonetic representation, not rehearsal per se.

Span length. Another expected manifestation of inefficient coding would be a reduced memory span. It is very possible that poor readers exhaust relatively more of their central processing capacity on the task of coding the

items and have a reduced recall span as a consequence (see Perfetti & Lesgold, in press). It was indeed the case that in each of the experiments, the reading groups differed in overall accuracy of recall. Our results are in agreement in this respect with earlier work by Naidoo (1970) and Miles and Miles (1977) in finding that reading ability is related to memory span in ordered recall.

We interpret the relatively briefer memory span of poor readers as the result of some deficiency in the use of phonetic coding. An alternative interpretation would treat the difference in memory span as the fundamental difference between good and poor readers, and would attribute the statistical interaction between reading group and phonetic confusability-nonconfusability to the greater difficulty of both the rhyming and nonrhyming tasks for the poor readers. The poor readers' limited span places them at or near chance level in the later serial positions on the more difficult task of recalling the rhyming items (see Figures 2 and 3) and therefore gives them less room to show an effect of phonetic confusability. There is no way to choose between these interpretations within the confines of the serial recall experiment. However, the investigation by Mark et al., 1977, to which we referred, demonstrated an unequivocal interaction between phonetic confusability and level of reading ability, but on a recognition memory task lacking the methodological difficulties inherent in the serial recall type of experiment.

Also relevant to the interpretation of our findings is the fact that poor readers, though impaired on tasks involving verbal material, may perform at the same level as good readers on nonlinguistic memory tasks (Vellutino, Pruzek, Steger, & Meshoulam, 1973; Vellutino, Steger, & Kandel, 1972). Two studies that find deficits in poor readers in recall of abstract figural patterns (Morrison, Giordani, & Nagy, 1977) and in recall of spatio-temporal patterns (Corkin, 1975) cannot properly be regarded as contradictory, since in both cases the tasks lend themselves to verbal labeling. Evidence from our own laboratory shows no significant differences between good and poor readers on a memory task employing highly abstract nonsense figures and faces (Liberman, I. Y., Mark, & Shankweiler, 1978).⁴ The existing data are consistent with the hypothesis that the deficiency of poor readers on memory tasks is limited to situations in which speech coding can readily occur.

New Directions

The preceding discussion suggests that the hypothesis of differences in the use of speech coding in working memory by good and poor readers may bring a unifying perspective to other often-cited difficulties of poor readers: Limited span in verbal recall and inefficient rehearsal. It remains to consider the consequences of the temporal order requirement of the task and to probe the origins of the phonetic coding deficit.

In view of suggestions in the literature (Crowder, 1978) that a major function of phonetic coding in working memory is to preserve information about temporal order, it is appropriate to consider whether difficulty specific to recalling the order of items is a manifestation of a faulty phonetic representation. With this possibility in mind, we rescored the subjects' responses in each experiment ignoring order and giving credit for any correctly recalled item regardless of the order in which it was written down.

The change in scoring procedure did not significantly alter the differences among the reading groups with regard to susceptibility to phonetic confusion. The present study, however, was not designed to distinguish order memory from item memory and therefore does not permit us to draw any definite conclusions as to whether good and poor readers differ in this respect. We are currently investigating the possibility that difficulties in ordered recall and recognition in poor readers are limited to situations in which speech coding is likely to occur. The question is the more interesting in view of Bakker's (1972) claim that in tests of perception and retention of information about order of items, the verbal or nonverbal nature of the task requirements is crucial.

It remains to be explored whether the problem that poor readers have in dealing with the phonetic representation stems from faulty establishment of phonetic encoding or reflects a difficulty of access to it. If the problem is chiefly of the latter kind, it will be important to discover what it is that limits access to the phonetic representation in poor readers. As for the hypothesis that the quality of the phonetic representation is the distinguishing factor, the possibility needs examination that subtle deficits might be demonstrated by children with reading disabilities in their perception of the acoustic cues for speech. Initially, this possibility seemed unlikely. There were no apparent difficulties in speech production or speech understanding in the poor reading groups. Indeed, these children were apparently indistinguishable from the superior readers. However, it is conceivable that, although there were no clinically apparent deficits in spoken language, suitably subtle analytic techniques, such as those used in the study of the acoustic cues for speech perception (Liberman, A. M., Cooper, Shankweiler, & Studdert-Kennedy, 1967) might reveal differences between the good and poor readers of this study.

Whether the origin of the language deficits in poor readers is in phonetic perception or whether it is specific to the memorial aspects of language, we may appropriately ask whether good and poor readers differ in susceptibility to phonetic confusions in memory for materials that are more like text designed for normal reading than are random letter strings. If poor readers typically have a genuine problem in phonetic coding, the effects should be demonstrable in sentence processing. At present, we are investigating differences between good and poor readers in recall of semantically meaningful and nonmeaningful sentences and random word strings, in which, for each type of material, a parallel comparison can be made between items that do and those that do not offer the opportunity for phonetic confusions to occur (Mann, Liberman, I. Y., & Shankweiler, Note 1).

Up to this point we have not considered possible alternatives to phonetic coding in working memory and their use in reading. The obvious possibility is that children with reading disability have a tendency to code memory representations of print into some visual or semantic form, and for that reason show relatively little susceptibility to phonetic interference. Conrad (1972) found that children younger than about the age of six typically employ a nonphonetic strategy in recall of pictured objects. He suggests that phonetic coding may not be available as a memory strategy for visual material in younger children, since, at about six, the normal children in his sample--but not the congenitally deaf taught by the manual method--spontaneously abandoned

pictorial coding in favor of phonetic coding. The problem exposed by Conrad of the development of working memory codes merits further study. In view of the present findings, showing closely parallel effects of phonetic similarity on recall of material presented visually and auditorily, it would be of interest to find out whether a comparable developmental shift in coding strategy occurs in normal children for recall of material presented by ear. We would expect, in any case, that individual differences in the age at which memory coding changes to a phonetically-based strategy would have a bearing on readiness to read.

Summary

The findings showed that poor readers make less effective use than good readers of a phonetic recall strategy in memory for letter strings. This result lends support to the hypothesis that differences in the use of phonetically organized representations in working memory are a relevant factor in learning to read. The poor readers' low susceptibility to phonetic interference in recall of rhyming letter strings may be due either to the unavailability of the phonetic representation to ready access, or to the degraded quality of such representations. Failure to make effective use of phonetic coding in memory was not limited to situations in which the materials were presented visually, but was manifested on auditory presentation as well. The poor readers' problem can therefore not be understood solely as a deficit in recoding from print, but as a more general deficiency in coding strategy, which may be expected to have consequences that extend beyond reading.

REFERENCE NOTES

1. Mann, V. A., Liberman, I. Y., & Shankweiler, D. Word sequences and sentence recall among good and poor readers. To appear in Haskins Laboratories Status Report on Speech Research, Haskins Laboratories, New Haven, Conn.
2. Wolford, G., & Porter, G. Simultaneous and sequential presentation of visual arrays. Paper presented at the annual meeting of the Psychonomic Society, November, 1976.

REFERENCES

- Baddeley, A. D. Short-term memory for word sequences as a function of acoustic and formal similarity. Quarterly Journal of Experimental Psychology, 1966, 18, 362-365.
- Baddeley, A. D. The trouble with levels: A reexamination of Craik and Lockhart's framework for memory research. Psychological Review, 1978, 85, 139-152.
- Baddeley, A. D., & Hitch, G. Working memory. In G. A. Bower (Ed.), The psychology of learning and motivation (Vol. VIII). New York: Academic Press, 1974.
- Bakker, D. J. Temporal order in disturbed reading. Rotterdam, The Netherlands: Rotterdam University, 1972.
- Bauer, R. H. Memory processes in children with learning disabilities: Evidence for deficient rehearsal. Journal of Experimental Child

- Psychology, 1977, 24, 415-430.
- Conrad, R. Acoustic confusions in immediate memory. British Journal of Psychology, 1964, 55, 75-84.
- Conrad, R. Speech and reading. In J. F. Kavanagh & I. G. Mattingly (Eds.), Language by ear and by eye: The relationships between speech and reading. Cambridge, Mass.: The MIT Press, 1972.
- Conrad, R., & Null, A. J. Input modality, acoustic confusion and memory span. British Journal of Psychology, 1964, 55, 429-432.
- Cooper, F. S., & Mattingly, I. G. Computer controlled PCM system for investigation of dichotic speech perception. Journal of the Acoustical Society of America, 1968, 46, 115. (Abstract)
- Corkin, S. Serial-ordering deficits in inferior readers. Neuropsychologia, 1975, 12, 347-354.
- Crowder, R. C. Language and memory. In J. F. Kavanagh & W. Strange (Eds.), Speech and language in the laboratory, school, and clinic. Cambridge, Mass.: MIT Press, 1978.
- Dunn, L. M. Peabody picture vocabulary test. Circle Pines, Mn.: American Guidance Service, 1965.
- Eriksen, D., Mattingly, I. G., & Turvey, M. T. Phonetic activity in reading: An experiment with kanji. Language and Speech, 1977, 20, 384-403.
- Fowler, C. A., & Liberman, I. Y., & Shankweiler, D. On interpreting the error pattern in beginning reading. Language and Speech, 1977, 20, 162-173.
- Frumkin, B., & Anisfeld, M. Semantic and surface codes in the memory of deaf children. Cognitive Psychology, 1977, 4, 475-493.
- Hintzman, D. L. Articulatory coding in short-term memory. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 312-316.
- Jastak, J., Bijou, S. W., & Jastak, S. R. Wide range achievement test. Wilmington, Del.: Guidance Associates, 1965.
- Kleiman, G. M. Speech recoding in reading. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 323-339.
- Levy, B. A. Reading: Speech and meaning processes. Journal of Verbal Learning and Verbal Behavior, 1977, 16, 623-638.
- Liberman, A. M., Cooper, F. S., Shankweiler, D., & Studdert-Kennedy, M. Perception of the speech code. Psychological Review, 1967, 74, 431-461.
- Liberman, I. Y., Mark, L. S., & Shankweiler, D. Reading disability: Methodological problems in information-processing analysis. Letter to the editor. Science, 1978, 200(4343), 801-802.
- Liberman, I. Y., Shankweiler, D., Liberman, A. M., Fowler, C., & Fischer, F. W. Phonetic segmentation and recoding in the beginning reader. In A. S. Reber & D. Scarborough (Eds.), Toward a psychology of reading: The proceedings of the CUNY Conference. Hillsdale, N.J.: Erlbaum, 1977.
- Locke, J. L. Phonemic effects in the silent reading of hearing and deaf children. Cognition, 1978, 6, 175-187.
- Mark, L. S., Shankweiler, D., Liberman, I. Y., & Fowler, C. A. Phonetic recoding and reading difficulty in beginning readers. Memory & Cognition, 1977, 5, 623-629.
- Miles, N. C., & Miles, T. R. Dyslexia as a limitation in the ability to process information. Bulletin of the Orton Society, 1977, 27, 72-81.
- Morrison, F. J., Giordani, B., & Nagy, J. Reading disability: An information processing analysis. Science, 1977, 196, 77-79.
- Naidoo, S. The assessment of dyslexic children. In A. W. Franklin & S. Naidoo (Eds.), Assessment and teaching of dyslexic children. London: Invalid Children's Aid Association, 1970.

- Perfetti, C. A., & Lesgold, A. M. Coding and comprehension in the skilled reader and implications for reading instruction. In L. Resnick & P. Weaver (Eds.), Theory and practice of early reading. Hillsdale, N.J.: Erlbaum, in press.
- Shankweiler, D., & Liberman, I. Y. Misreading: A search for causes. In J. F. Kavanagh & I. G. Mattingly (Eds.), Language by ear and by eye: The relationship between speech and reading. Cambridge, Mass.: MIT Press, 1972.
- Shankweiler, D., & Liberman, I. Y. Exploring the relations between reading and speech. In R. M. Knights & K. Bakker (Eds.), The neuropsychology of learning disorders: Theoretical approaches. Baltimore: University Park Press, 1976.
- Swisher, L. The language performance of the oral deaf. In H. Whitaker & H. A. Whitaker (Eds.), Studies in neurolinguistics (Vol. II). New York: Academic Press, 1976.
- Tzeng, O. J. L., Hung, D. L., & Wang, W. S-Y. Speech recoding in reading Chinese characters. Journal of Experimental Psychology: Human Learning and Memory, 1977, 3, 621-630.
- Vellutino, F. R., Pruzek, R. M., Steger, J. A., & Meshoulam, U. Immediate visual recall in poor and normal readers as a function of orthographic-linguistic familiarity. Cortex, 1973, 9, 368-384.
- Vellutino, F. R., Steger, J. A., & Kandel, G. Reading disability: An investigation of the perceptual deficit hypothesis. Cortex, 1972, 8, 106-118.
- Wolford, G., & Hollingsworth, S. Evidence that short-term memory is not the limiting factor in the tachistoscopic full-report procedure. Memory and Cognition, 1974, 2, 296-300.

FOOTNOTES

¹In considering the role of phonetic short-term memory in reading, we make no assumptions about the possible role that speech-related processes might play in word recognition. We would mention, however, that the reader of an alphabetically written language must derive a phonological representation from the orthography if he is to gain a major advantage of alphabetic writing: namely, the possibility of decoding new words never before seen in print. The mode or modes of lexical access, in the case of familiar words, is, of course, a separate question, and one that is not relevant to our concerns in this paper. The need to distinguish the possible role of speech coding in lexical access from its role in working memory for stretches of text longer than the word is underscored by the findings, to which we referred, on readers of Chinese and Japanese. Users of these logographic orthographies might or might not enter the internal lexicon via a phonological representation; that is an open question. What is clear, however, from the findings of Tzeng, Hung, and Wang (1977) and Erickson, Mattingly, and Turvey (1977) is that these logographic readers, like most adult readers of English, make predominantly phonetic confusions when they attempt to hold strings of logograms in short-term memory.

²Preliminary and incomplete accounts of portions of the findings presented here were included in I. Y. Liberman et al. (1977) and in Shankweiler and I. Y. Liberman (1976).

³The question arises whether the phonetic and visual characteristics of the letter strings might have been confounded, with the effect of obscuring the interpretation of the results. In order to assess phonetic confusability independent of any confounding effects of visual similarity, we carried out an additional analysis of the data of Experiments 2 and 3, examining only the errors in which phonetic confusion, but not visual confusion, could be implicated (e.g., "B" occurred as the response at the position in which Z occurred in the stimulus string). Thus, this analysis excluded from consideration errors that are ambiguous (e.g., the response "B" for P). Classification of the errors was based on the results of visual and phonetic similarity scaling by Wolford and his colleagues (Wolford & Hollingsworth, 1974; Wolford & Porter, Note 2). The results of the analysis of unambiguous cases showed that good readers still uniformly made a significantly higher proportion of phonetic confusions than the poor readers. Details of this analysis will be made available to the reader upon request.

⁴A full account of this study, which includes M. Werfelman as a co-author, is in preparation.

SOME EXPERIMENTS ON THE SOUND OF SILENCE IN PHONETIC PERCEPTION*

Michael F. Dorman,⁺ Lawrence J. Raphael⁺⁺ and Alvin M. Liberman⁺⁺⁺

Abstract. The results of several experiments demonstrate that silence is an important cue for the perception of stop-consonant and affricate manner. In some circumstances, silence is necessary; in others, it is sufficient. But silence is not the only cue to these manners. There are other cues that are more or less equivalent in their perceptual effects, though they are quite different acoustically. Finally, silence is effective as a cue when it is part of an utterance that is perceived as having been produced by a single male speaker, but not when it separates utterances produced by male and female speakers. These findings are taken to imply that, in these instances, perception is constrained as if by some abstract conception of what vocal tracts do when they make linguistically significant gestures.

INTRODUCTION

The several experiments to be reported here have in common a concern with silence as one of the cues for the perception of stop consonants. They were designed to illuminate further the processes by which that cue does its perceptual work.

That silence is important for the perception of stops has been established by several studies. Indeed, silence has been found to play a role in perceiving each of the three features--manner, voicing and place--that a stop consonant comprises. Consider manner. By cutting and splicing magnetic

*The results of Experiments I, II and VI were described in a paper presented at the 89th Meeting of the Acoustical Society of America, Austin, Texas, 1975; Experiments III, IV and VII at the 91st Meeting of the Acoustical Society of America, Washington, D.C., 1976, and the results of Experiments Va and Vb at the 93rd Meeting of the Acoustical Society of America, State College, Pennsylvania, 1977.

⁺Also University of Arizona, Tempe.

⁺⁺Also Herman H. Lehman College and The Graduate School, City University of New York.

⁺⁺⁺Also Yale University and University of Connecticut, Storrs.

Acknowledgment: This research was supported by a grant (HD-01994) from the National Institute of Child Health and Human Development to Haskins Laboratories. We wish to thank Bruno Repp for helpful comments on an earlier version of this paper. We also wish to thank Anthony Levas and Suzi Pollack for their assistance in collecting and tabulating portions of the data.

tapes, Bastian, Rimas, and Liberman (1961) showed that the syllable "slit" is heard as "split" when a short interval of silence (about 40 msec) is introduced between the noise at the beginning of the syllable and the vocalic portion. As for voicing, Lisker (1957a) early found that intervocalic stops in trochees were perceived as voiced or voiceless (for example, "rabid" or "rapid") depending on the duration of silence between the syllables. Turning finally to place, we take account of the finding by Port (1976) that "rabid" is perceived as "ratted" when the duration of silence between the syllables is reduced.

Our experiments will deal only with the perception of stop-consonant manner. Taken together, and added (when appropriate) to the work of others, they are meant to bear on three related questions: (1) In what circumstances is silence a cue? (2) Does silence have its effect exclusively in the auditory domain, or also at some more abstract (phonetic) remove where perception is constrained as if by knowledge of what a vocal tract does when it makes linguistically significant gestures? (3) If the latter, then whose vocal tract provides the constraint?

SILENCE AS A NECESSARY CONDITION BEFORE AND AFTER THE VOWEL; PERCEPTION OF TRANSITION CUES IN SPEECH AND NONSPEECH CONTEXTS

Evidence pointing to the importance of silence as a manner cue came first from experience with syllables in which a stop is (or is not) heard before the vocalic nucleus. Thus, in the early study by Bastian et al. (1961), the contrast was between "slit" and "split." Given similar phonetic contexts, the same effect is readily found, so readily indeed that it has become part of the lore of those who experiment with speech, and is taken into account in formal rules that specify how speech is to be synthesized. In contrast, there is little information about the importance of silence as a manner cue for the perception of stops that follow the vocalic nucleus. We can infer, however, from an early observation by Lisker (1957a) and a more recent study by Abbs (1971) that a silent interval of some length must follow a vowel-stop syllable if the stop is to be perceived.

Our aim is to learn more about these phenomena. To that end, we will first assess the role of silence in the perception of stops (before the vowel) in the syllables [pɪ] and [kɪ] and (after the vowel) in the disyllables [bɪdɪ], [bɪg dɪ] and [bɪd dɪ]. If, as we have reason to expect, silence proves to be important, we will use the results as a basis for further studies that might help us to understand why. Some of those will be reported in this section, others in the sections that follow.

To see what choices we face when we wonder why silence should be a cue for stops, we should first consider the perceptual consequences of altering the acoustic structure of the fricative-vowel syllable shown in Figure 1: having recorded a naturally produced token of [sa], we find that removing the initial fricative noise will often leave a syllable that sounds like [da]; if we store the noise, but move it backward in time so as to leave a brief (say 50 msec) interval of silence between it and the vocalic portion of the syllable, we produce a syllable that sounds like [sta] (Bastian, 1962). At one level of interpretation there is no mystery in this: The fricative [s] and the stop [t] have similar places of production, hence similar formant

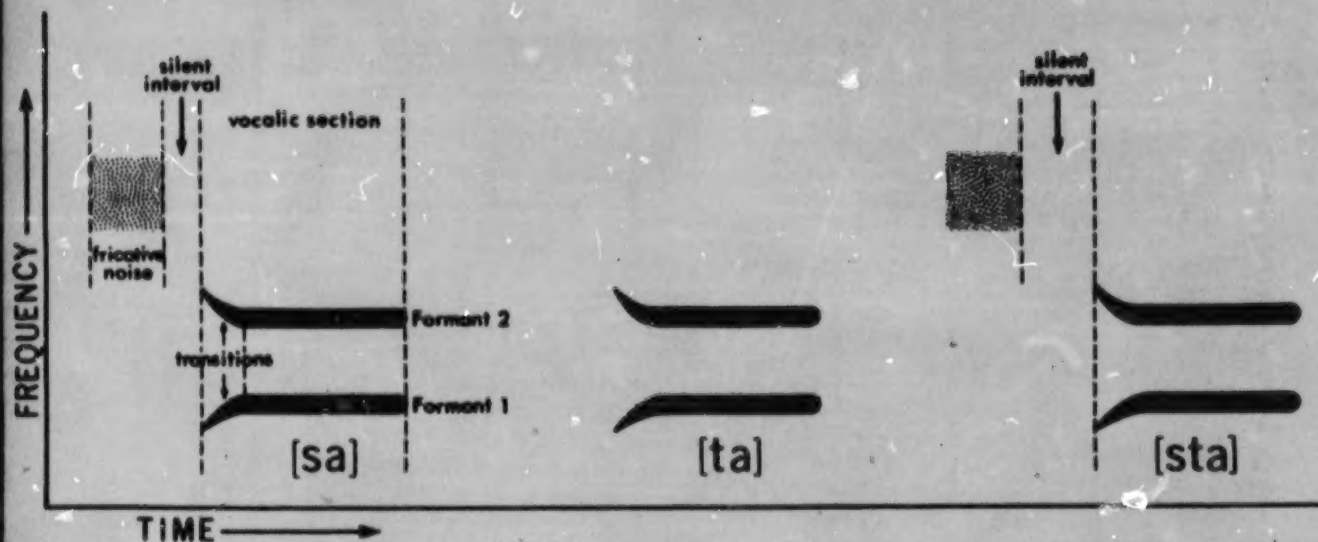


Figure 1. Schematic representation of stimulus patterns sufficient for the perception of [sa], [ta] and [sta]. Adapted from Liberman and Pisoni, 1977.

transitions. However, it is not so clear why silence is necessary in order for the transition cues to give rise to the perception of a stop--that is, why a stop is not heard when fricative noise and formant transitions are separated by only a brief interval.

Broadly speaking, two interpretations are possible. The one we are inclined to favor is that the silence provides information to a (phonetic) perceiving device that is specialized to make appropriate use of it. To see why that is at least plausible, consider that a speaker cannot produce a stop without closing his vocal tract, and that he cannot close his vocal tract without producing a corresponding period of silence. When the listener hears an insufficiently long period of silence between the fricative noise and the vocalic section, it is, by this account, as if he "knew" that a stop should not be perceived because it was not produced.

An alternative interpretation puts the effect of the silence cue squarely in the auditory domain. Thus, we note about the example just offered that it conforms to the paradigm for auditory forward masking. Conceivably, the fricative noise masks the transition cues that otherwise would be sufficient for the stops; in that case, the role of silence would be to provide time to evade masking. Or, keeping the interpretation still in the auditory domain, we might suppose that the silence collaborates in some kind of perceptual interaction with the transition cues, the result of the interaction being that experience we call a stop.

Some evidence relevant to these interpretations is already available. Harris (1958), for example, found recognition of the [f] - [θ] contrast to be contingent primarily on the formant transitions that follow the fricative noise. This situation could only arise if the formant transitions had different effects in the auditory domain--that is, if they were not masked by the preceding noise. Evidence from dichotic listening supports this conclusion. Thus, Darwin (1971) found a larger right ear advantage (REA) for fricatives synthesized with appropriate formant transitions following the fricative noise than for fricatives synthesized without formant transitions. In this instance, too, the transitions must have had different auditory representations when they arrived at the central processing mechanisms responsible for the ear advantage.

Another piece of relevant evidence comes from a study of selective adaptation. Following a now standard adaptation procedure, Ganong (1975) first measured the displacement of the [b₁ - d₁] boundary caused by adaptation with [d₁]. Fricative noise was then placed in front of the [d₁] and the (perceived) [s₁] that resulted was used as the adapting stimulus. The outcome was a shift in the [b₁ - d₁] boundary as large as that found when the adapting stimulus was [d₁]. Patterns that contained the noise but not the formant transitions did not produce as large a shift. This indicates not only that the transition cues were getting through, but that they were getting through in full strength.

Thus, we are led to believe that the transition cues make a significant perceptual contribution, whether or not they are preceded by a period of silence. On that view, silence is important, not because it provides time to evade masking, or because it collaborates in an auditory interaction, but

because it provides information that is essential in determining how the transitions are to be interpreted in phonetic perception.

The experiments in this section are designed to get at that matter via a different--perhaps more direct--route by comparing the effect of the fricative noise on transition cues that are, in one case, in a speech context and, in the other, not. The results will bear, of course, on a masking interpretation, but also on the possibility of auditory interactions, since we will be able to determine whether or not there are qualitative changes in the perception of the nonspeech transition cues depending on the presence or absence of the silence.

EXPERIMENT I

Our first experiment was designed: (1) to assess the role of silence in the perception of stop manner prevocally in the syllables [pɪ] and [kɪ], and (2) to determine whether the fricative noise of [ʃ] masks or interacts with information carried on the transition cues for the stops when those are isolated from the rest of the syllable and are heard as nonspeech.

Method

Two sets of stimuli were made. Members of the one set--to be referred to as the "speech" stimuli--were appropriate for determining the effect of silence on the perception of the stop consonants in [pɪ] and [kɪ]. They were made in the following way. First, the syllables [ʃɪ], [gɪ], and [bɪ] were recorded by a male speaker, then digitized and stored, using the Pulse Code Modulation (PCM) system at Haskins Laboratories.¹ Working from high-resolution oscillograms and taking advantage of computer control, we next separated the fricative noise of the [ʃ] from the vocalic portion of the syllable [ʃɪ] and removed the syllable-initial bursts from the [gɪ] and [bɪ]. To create the experimental stimuli, we prefixed the ʃ-noise to what remained of the [bɪ] and [gɪ], leaving silent intervals of 0, 4, 8, 12, 16, 20, 40, 60, 80, and 100 msec between the offset of the fricative noise and the vocalic section appropriate for [gɪ] and [bɪ] (see Figure 2a for a schematic representation of one of the ʃ-noise plus [gɪ] stimuli). Four tokens of each stimulus type were produced. These were randomized and recorded on magnetic tape with a three-sec interval between stimuli.

Members of the other set--to be referred to as the "nonspeech" stimuli--were intended to enable us to measure the extent to which the transition cues that distinguish the stops in [pɪ] and [kɪ] are themselves masked by the ʃ-noise. These stimuli were made in the following way. First, the [bɪ] and [gɪ] patterns of the speech set were band-pass filtered between 0.9 and 3.5 kHz and truncated so as to include only the first 50 msec of the signal. This procedure eliminated the first formant, producing signals that contained only the second- and third-formant transitions. (Listeners could hear these stimuli as "chirps," and we supposed that with only a few minutes of practice they would be able to identify them by pitch as "low" or high.) Then, to create a test of the identifiability of these transitions for comparison with the condition in which they were the essential cues for place of articulation, we prefixed the ʃ-noise, setting the same intervals of silence between it and the chirps that we had used in creating the "speech" stimuli. (See Figure 2b for

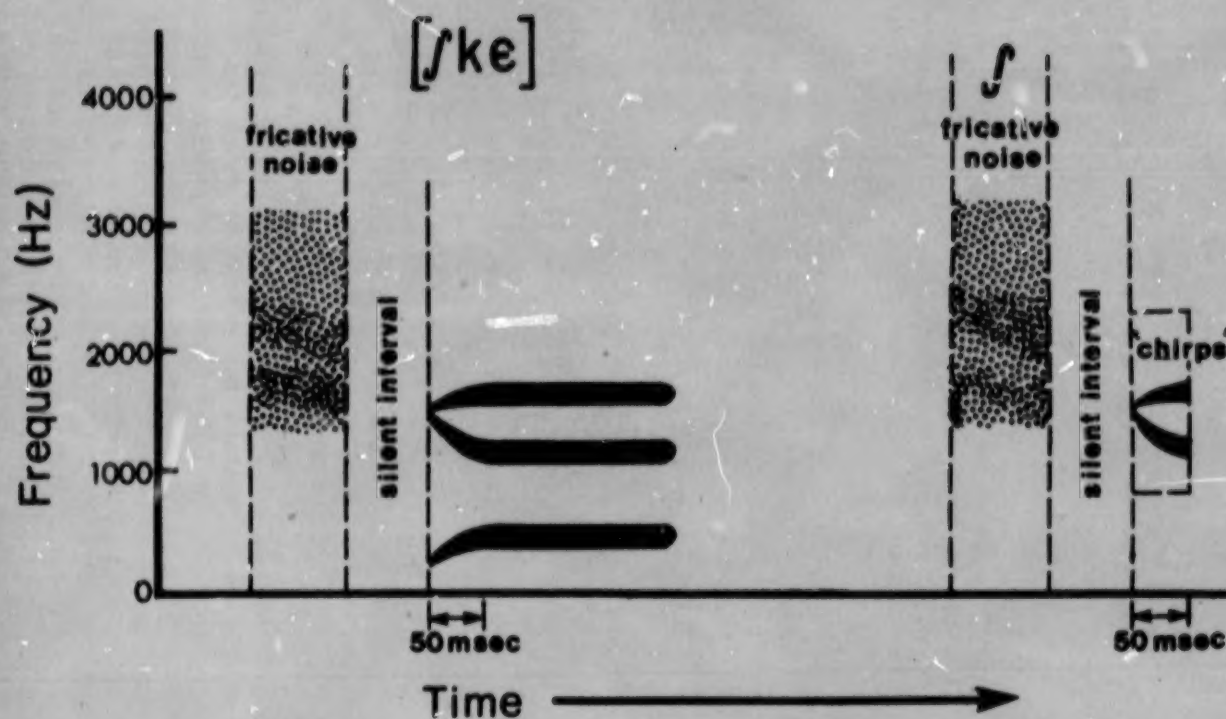


Figure 2. (a) Schematic representation of one of the speech patterns used in Experiment I.

Figure 2. (b) Schematic representation of the corresponding nonspeech ("chirp") pattern.

a schematic representation of the "chirp" stimulus derived from the "speech" stimulus shown in Figure 2a.) The resulting signals were randomized and recorded on magnetic tape with a three-sec interval between stimuli.

The subjects were nine volunteers, all undergraduates at Lehman College, who had not previously served in experiments on speech perception. Divided into groups of five and four, they listened in a sound-attenuated room, first to the speech stimuli and then, in a second session, to the "nonspeech" stimuli. In the speech condition, the listeners were told they would hear approximations of the syllables [ʃpi], [ʃki] and [ʃi] and were asked to indicate on a printed response sheet what they had heard. To provide some "practice," we presented twenty of the stimuli before the experiment proper began; no information was given about the "correctness" of the responses.

In the "nonspeech" condition, the subjects were told they would hear tokens of three stimulus types: ʃ-noise alone, ʃ-noise followed by a low-pitched chirp (which they were to call "low"), or ʃ-noise followed by a high-pitched chirp (which they were to call "high"). They were asked to indicate on their response sheets what they had heard. In this condition, the "practice" consisted of presenting 50 of the stimuli. In order to make sure that the subjects did, in fact, learn to identify the chirps, we provided knowledge of results. To preclude biasing the experimental outcome by experience during the practice sessions, we avoided all short silent intervals--in which the chirps might or might not be heard--presenting only those stimuli in which the noise preceded the chirps by 100 msec. During the experimental session, no information about "correct" responses was given.

In both "speech" and "nonspeech" conditions the stimuli were reproduced via a Revox 1240 tape recorder and AR-4x loudspeaker.

Results and Discussion

The results for the speech condition are shown in Figure 3. Since the identification functions for [ʃpi] and [ʃki] were found on preliminary examination to have similar shapes, we have averaged them; this facilitates comparison with the identification function for [ʃi]. We see that when the silent interval was less than 20 msec, listeners reported hearing [ʃi]--that is to say they did not hear a stop. The stops were identified with 75 percent accuracy only when the silent interval exceeded about 40 msec. Thus, we find silence to be an important condition for the perception of stops in fricative-stop-vowel syllables.

The identification functions shown in Figure 3 were derived from the responses of seven of the nine subjects. The two other subjects identified the ʃ-noise plus [gi] stimuli in the same manner as the group of seven, but made a total of only one [ʃi] response to the ʃ-noise plus [bi] stimuli. To account for that we should consider that in the case of [ʃpi], the places of articulation signaled by the fricative noise and the vocalic transitions were quite different, the former being palatal and the latter bilabial. In our own listening to these patterns, it seemed that when there was little silence between ʃ-noise and [bi], we heard [ʃi]; but with a nonspeech chirp--as if the transitions could not be integrated into the phonetic percept but were audible nevertheless. It is possible that our subjects, hearing the same chirp,

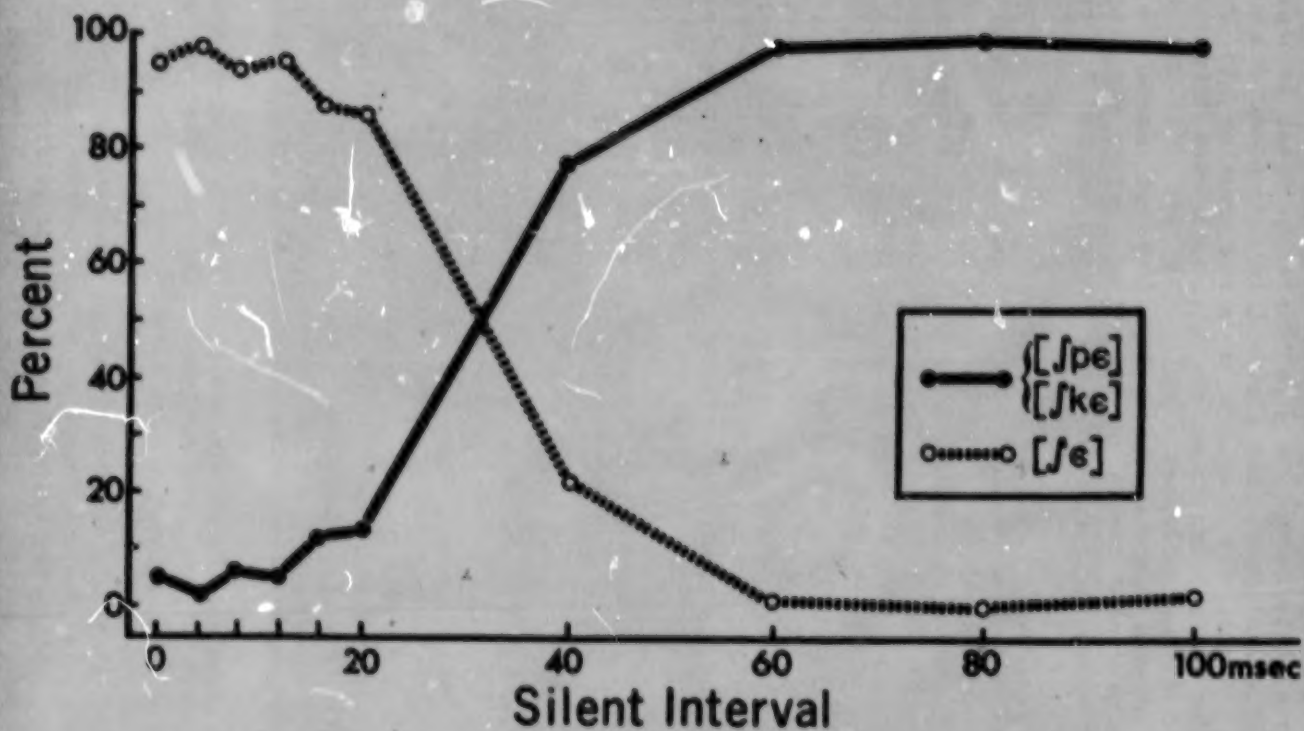


Figure 3. Silence as a necessary condition for stop manner; identification of stimulus patterns as {fpe}-[fke] or [fɛ].

elected to call these stimuli [ʃpə]. In the case of ʃ-noise plus [gə], the disparity in place of articulation was not so great, and it is perhaps for that reason that when the -noise was moved close to the [gə] we, and all our subjects, heard only [ʃə]. Indeed, the disparity in place of articulation can be reduced even further, as it is, for example, in the case of s-noise plus [tə] that we described in the introduction. There, the places of articulation for the fricative and stop are exactly the same, and the [sə] that results from putting the fricative noise close to the vocalic section is virtually indistinguishable from one that is produced by a human speaker who articulates in a perfectly normal way.

We should emphasize that the interval of silence necessary for stop perception in fricative-stop-vowel syllables is not invariant. Indeed, from the early work of Bastian and from recent work by Bailey, Summerfield, and Dorman (Note 1) and by Summerfield and Bailey (1977), we know that the interval varies according to how several other cues are set. These include, at the least, the duration of the fricative noise, the rate of fricative noise offset, the rise-time of the amplitude envelope of the vocalic portion of the syllable, and the starting frequency of the first-formant transition. (We discuss the importance of such relations among cues more fully in the next section.)

We should also emphasize that we do not mean to imply that listeners cannot discriminate between a naturally produced [ʃə] and one composed of ʃ-noise followed at a brief interval by [gə] (or [bə]). As we pointed out above, in these cases a listener may hear a normal [ʃə] or [ʃə] with a nonspeech chirp in it. Now we should add that for some articulations of [gə], a fricative noise placed just in front will cause a listener to perceive [ʃə] (Liberman & Pisoni, 1977). The point we wish to make is that listeners do not in such cases commonly report a stop.

Redirecting our attention to Experiment I, we see in Figure 4 that the results of the nonspeech condition are quite different from those of the speech condition. The isolated formant transitions taken from [bə] and [gə] were clearly audible--indeed, highly identifiable--as chirps at all intervals of silence, even zero. That outcome is wholly consistent with the evidence presented at the introduction to this section in that transition cues that follow fricative noise are nonetheless effective as auditory events, whether separated from the noise or not. As for the possibility that the transition cues somehow interact with silence, there had previously been no data that were directly relevant. Now we see in the results of our experiment a suggestion that such auditory interaction does not occur: Our subjects not only heard the nonspeech transitions (no matter how close they were to the fricative noise), but they correctly identified them as well; moreover, our own listening made it plain that, more generally, the fricative noise did not appreciably affect the perception of the nonspeech transitions in any qualitative way.

EXPERIMENTS IIA AND IIB

In the previous experiment we found silence to be a necessary condition for the perception of stops in prevocalic position. The experiments reported here were designed to find out if silence is also a necessary condition for

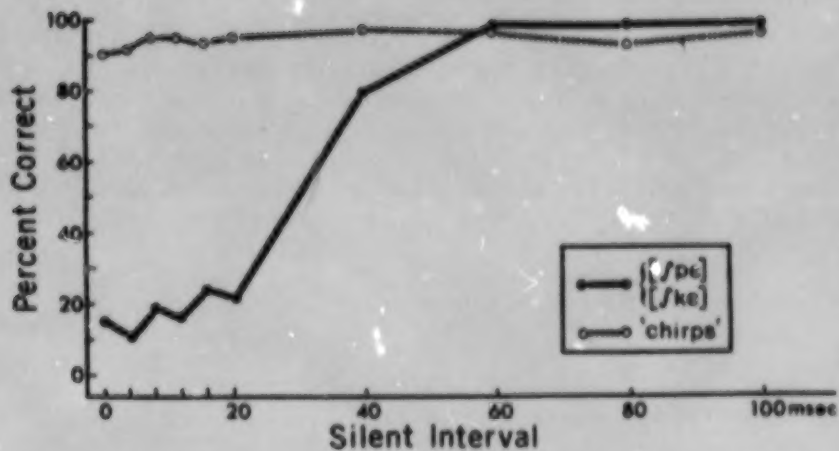


Figure 4. Percent correct identification of the transition cues in the speech ([pt]-[kt]) and nonspeech (chirps) contexts.

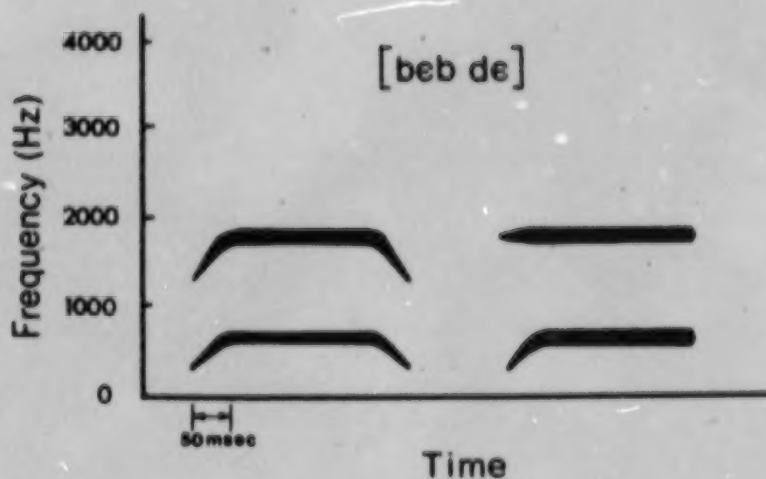


Figure 5. Schematic representation of one of the stimulus patterns for Experiment IIa.

the perception of stops in postvocalic position. There were two such experiments, divided according to purpose and the nature of the stimuli.

In one experiment (IIa), the stimuli were the synthetic disyllables [bɑb dɑ] and [bɑg dɑ], made to provide variation in the interval of silence between the first and second syllables. Given the hypothesis that underlies all the experiments of this paper, we should expect that a relatively long silence would be essential if the listener is to perceive both the syllable final [b] and [g] and the syllable initial [d], since a speaker must close his vocal tract for a longer period to say [bɑb dɑ] or [bɑg dɑ] than to say [bɑ dɑ], [bɑ bɑ] or [bɑ gɑ]. Pilot work revealed that with reductions in the duration of the silent interval, it was the syllable-final stops [b] and [g] that disappeared; the syllable-initial [d] could be heard even at very short intervals of silence. This may be owing, in part, to the fact that, in production, the [d], and especially the flapped [d], requires very little closure (Port, 1976), and in part, perhaps, to the fact that unreleased syllable-final stops tend to be relatively unintelligible. At all events, it is the syllable-final stops that are, in the kinds of patterns we used, the more sensitive to variations in the duration of intersyllabic silence.

As in the experiments with prevocalic stops, we thought it useful to provide data relevant to the possibility that the outcome is to be accounted for in terms of masking--backward masking in the case of the postvocalic stops--or auditory interaction. To that end, we determined whether silence is also necessary for the perception of the formant transitions that are sufficient to distinguish the syllable-final stops when those transitions are presented in isolation and sound like chirps.

In the other experiment (IIb), the stimuli were natural speech, not synthetic, and they included not only [bɑb dɑ] and [bɑg dɑ] but also the geminate condition [bɑd dɑ].² The use of natural speech will permit a comparison with the results obtained when the stimuli were synthetic. The point of testing the geminate condition is that, in production, the articulatory closure for the geminate stops is longer than that for single stops, and a study by Pickett and Decker (1960) leads us to suspect that the amount of silence necessary for perception may also be longer. A comparison of the two cases of syllable-final stops seemed, therefore, to be in order.

Method

To produce stimuli for Experiment IIa--the one with synthetic stimuli--we used the Haskins Laboratories parallel-resonance synthesizer to generate two-formant patterns appropriate for the disyllables [bɑb dɑ] and [bɑg dɑ]. A schematic representation of [bɑb dɑ] is shown in Figure 5. That disyllable differed from the other one [bɑg dɑ] in the second-formant transition, the sole cue in these patterns for the perceived distinction between the syllable-final stops: For [b] the transition is falling, as shown in the figure, while for [g] it is rising. We then introduced periods of silence between the second syllable [dɑ] and the first syllable [bɑb] or [bɑg]. These periods ranged from 0 to 150 msec in steps of 10 msec. Four tokens of each stimulus were generated. To produce a test sequence appropriate for presentation to our subjects, we put these stimuli into a random sequence with a three-sec interval between successive stimuli. That test sequence was used in what will

be referred to as the "speech" condition.

To produce the corresponding stimuli for the "nonspeech" condition, we simply isolated the second-formant transitions that alone distinguished the [bab] and [bag] patterns of the "speech" stimuli (falling for [b], rising for [g]), and then produced stimuli that were otherwise identical with those of the "speech" condition; that is, we placed after the isolated transitions the same synthetic [da] that had been used in the "speech" condition and introduced between it and the transitions the same intervals of silence.

The subjects for Experiment IIA were six undergraduates at Lehman College who had previously participated in experiments on speech perception. They were tested individually. Test order ("speech" vs "nonspeech") was counter-balanced across subjects. In the "speech" condition, the subjects were asked to respond [bab da], [bag da], or [ba da] and to write their responses. To familiarize the subjects with the stimuli, we had them listen to twenty of the patterns before the experiment began. The stimuli were reproduced on a Revox 1240 tape recorder via TDH 39 headphones.

In the "nonspeech" condition, the subjects were told they would hear a high-pitched chirp followed by [da], a low-pitched chirp followed by [da], or [da] alone. They were asked to respond accordingly. To teach the subjects to identify the chirps, and to make sure they could reliably do so, we first presented 50 [b] and [g] chirps in random order with feedback of results. Then we presented, also in random order, twenty-five [b] and [g] chirps followed in each case, after a 120-msec interval, by [da]. Again, subjects were told the correct answers after they had made their responses. The point of using only the 120-msec interval was to avoid biasing the results by providing "correct" responses in those cases where the [da] syllable was sufficiently close for "masking" to have conceivably occurred. Finally, the test proper was begun. During the test, there was, of course, no providing of "correct" answers.

The procedures for Experiment IIB, the one that included the geminate case and was done with natural speech, were as follows. Having recorded a male speaker saying [bab], [bad], [bag] and [da], we used the editing facilities provided by the Maskins Laboratories' PCM systems to truncate closure voicing following the syllable-final transitions to 15 msec. To each of the syllables [bab], [bad] and [bag], we then appended the syllable [da], separating it from [bab], [bad] or [bag] by periods of silence that ranged from 0 to 90 msec in steps of 10 msec. Three tokens of each stimulus were generated. These were randomized and recorded onto tape with a four-sec interval between stimuli.

The subjects for this experiment were eight volunteers, all undergraduates at Lehman College who had not previously served in speech-perception studies. They were asked to identify each of the stimuli as [bab da], [bag da], [bad da] or [ba da] and, in writing their responses, to include the entire syllable. There was a preliminary "practice" session in which the subjects heard and identified twenty stimulus patterns. The signals were produced in the manner described in Experiment I.

Results and Discussion

The effect of silence on the perception of syllable-final stops in synthetic [bɔb dɔ] and [bɔg dɔ] (Experiment IIa) is shown in Figure 6. There we have plotted the average [bɔb dɔ] and [bɔg dɔ] responses for comparison with the [bɔ dɔ] responses. (The identification functions for [bɔb dɔ] and [bɔg dɔ] were similar, so we have collapsed them into a single function.) One sees that, over the range 0 to about 30 msec of intersyllabic silence, the predominant response was [bɔ dɔ]; that is, our subjects did not report a syllable-final stop.³ We should emphasize that, as in the experiment on prevocalic stops, it was not the case that a subject heard a stop but misidentified it; rather, he simply did not hear it. A silent interval of about 58 msec was necessary before the subjects identified the stops with 75 percent accuracy. Thus, for the perception of stops in postvocalic position, as for those that were prevocalic, silence is important.

It will be remembered that we were also concerned with how the isolated formant transitions of the syllable-final [b] and [g] (nonspeech condition) are affected when the stimulus patterns are otherwise exactly the same as in the speech condition just reported. The results of the nonspeech condition are shown in Figure 7. We note, first, that no subject used the response "no chirp"; that is, no subject ever failed to hear a chirp, even when there was no silence between the chirp and the syllable. This is dramatically different from the result obtained in the "speech" condition. There, given comparable conditions, our subjects did not hear the corresponding syllable-final stops at all. Looking at the percentage of chirp identification, however, we see that at the shortest intervals of silence, identification is less accurate than at the longest intervals. Indeed, this difference in accuracy is significant, $F = 2.07$, $p < .05$. We should note, however, that even at the brief intervals our listeners averaged about 70 percent correct. Thus, it does not appear that backward masking can account for the complete absence of the stop percept at brief silent intervals.

We turn now to the results of Experiment IIb. This experiment differed from the previous one in that the geminate condition was included, and natural rather than synthetic speech was used. Let us first compare the results obtained with natural speech and with synthetic speech. For that purpose we will look only at the data pertaining to syllable-final [b] and [g], omitting the geminate condition. These are shown in Figure 8, together with the comparable data (from Figure 7) for synthetic speech. The results are quite similar; in both conditions some interval of silence is necessary for listeners to identify a stop. However, the duration of that interval does differ by about 15 msec between the two conditions. We should suppose that this difference is due to variation between the conditions in the "settings" of the cues (for stop manner) other than silence, for example, formant transitions.

Turning now to the comparison between geminate and nongeminate stops, we see in Figure 9 that subjects needed a longer silent interval to identify syllable-final [d] than [b] or [g];⁴ even at the longest interval the identification of [d] reached only 38 percent correct. Further research by Repp (1976) suggests that an interval of approximately 200 msec is necessary for listeners to identify the syllable-final stop in a sequence of identical

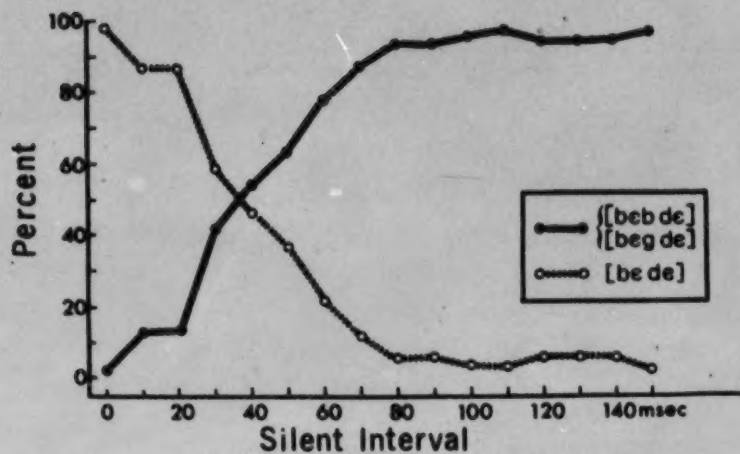


Figure 6. Silence as a necessary condition for stop manner; identification of stimulus patterns as [beb d_g]-[beg d_g] or [ba d_g].

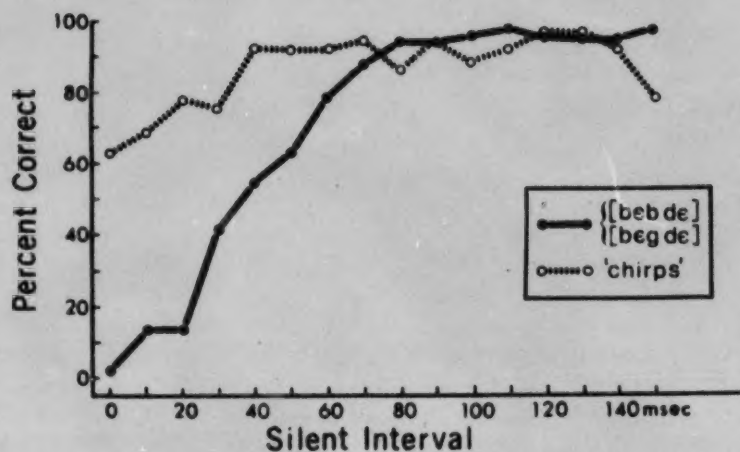


Figure 7. Percent correct identification of the transition cues in the speech ([beb d_g]-[beg d_g]) and nonspeech (chirps) contexts.

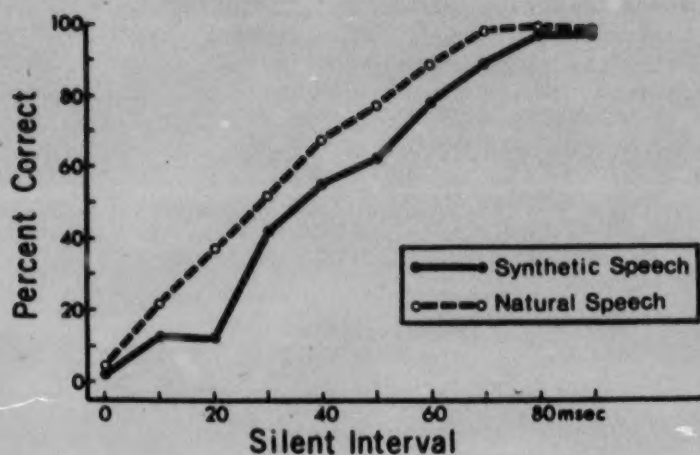


Figure 8. Identification functions for syllable-final stops in synthetic and natural speech.

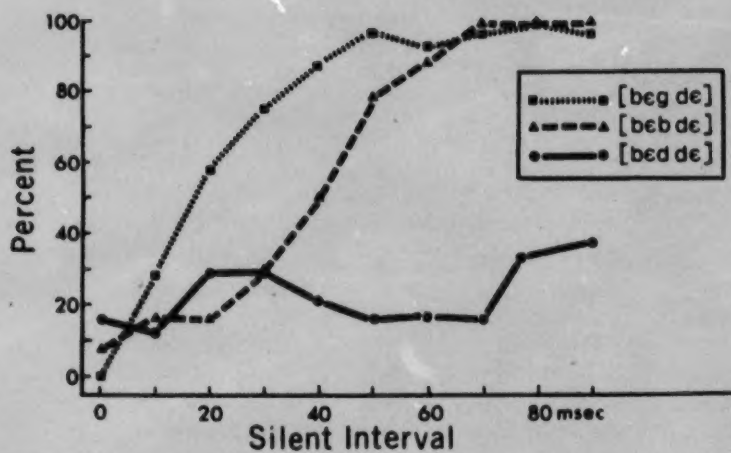


Figure 9. Identification of syllable-final stops in geminate ([bɛd dɛ]) and nongeminate ([bɛb dɛ] and [bɛg dɛ]) conditions.

stops (see also Pickett & Decker, 1960; Fujisaki, Nakamura, & Imoto, 1975). This result provides further evidence against an explanation of the perceptual disappearance of the syllable-final stops in terms of recognition masking, for one would be hard-pressed to explain why syllable-initial [d] should "backward-mask" syllable-final [d] over a period four times longer than it masks [b] or [g].

More direct evidence that syllable-final transitions are not "backward masked" is also to be found in studies by Repp (1976, 1977). Having presented to listeners VCV's that had been synthesized with and without syllable-final transitions, he found, in the case of stimuli without syllable-final transitions, that the time required to identify the medial consonant increased as a function of the duration of the closure interval; in the case of stimuli with syllable-final transitions, however, the time required was more nearly constant (Repp, 1976). Clearly, then, the syllable-final transitions had a perceptual effect even though they were not heard as discrete phonetic events. This same conclusion can be drawn from another experiment by Repp (1977). In that experiment the syllable [dɛ] was preceded, in the one case, by [ad], in the other case by [ab]. In both cases the listeners perceived [adɛ]. Nevertheless, they discriminated between the stimuli at a level slightly better than chance.

Returning now to our own results, we conclude from Experiments IIa and IIb that, just as silence is important for the perception of stops in prevocalic position, so also is it important for the perception of stops in postvocalic position. Moreover, the results are consistent with the evidence presented in the introduction--namely, that silence is important, not because it provides time to evade masking or because it enters into an auditory interaction, but rather because it provides information about the behavior of a vocal tract.

SILENCE AS A SUFFICIENT CONDITION BEFORE AND AFTER THE VOWEL: PERCEPTUAL EQUIVALENCE OF SILENCE AND SOUND

In the studies so far described, stops were (or were not) perceived in patterns that contained transition cues appropriate for stop manner. Now we shall turn to cases in which the transition cues are absent, and it is left to the power of the silence cue itself to produce the effect of a stop. We should note that even in the early study by Bastian et al. (1961), silence might have borne the entire burden, but we cannot be sure because the procedures of cutting and splicing the magnetic tape may have introduced a transient, which of itself could contribute to the perception of a stop. We should also note that others (Summerfield & Bailey, 1977), working independently of us, have recently demonstrated the power of silence to cue stop manner prevocalically in the context of fricative-vowel ya fricative-stop-vowel, for example, [si] ya [ski], where the vocalic section alone is, by perceptual test, not sufficient to produce the stop. At all events, we, too, wish to test the silence cue in such circumstances, and to do it for several positions in the syllable: in prevocalic position ("slit" ya "split"); in intervocalic position ("say shop" ya "say chop," the affricate "ch" [tʃ] being taken here as a stop-initiated fricative); and in postvocalic position ("dish" ya "ditch"). The results may throw more light on the role of silence in the perception of stop manner, since in these instances there are no obvious

transition cues to be masked. They will also provide the basis for further investigations into the reasons why silence should have a role in stop perception at all.

To see the point of one of these further investigations we should recall that the role of silence might be to tell the listener that the speaker either did or did not close his vocal tract appropriately for the production of a stop consonant. However, to make that suggestion is to imply that our perception of speech is constrained to some degree by a device that acts as if it knew what vocal tracts can and cannot do when they make linguistically relevant gestures; or, more generally, that there is, in speech, a link between perception and production. Further evidence for such a link comes, for example, from studies that have established an equivalence in phonetic perception between cues that are very different from an acoustic (and presumably auditory) point of view, but which are the correlated results of the same articulatory gesture. One of the earliest of these is of special interest to us because it dealt with silence, albeit as a cue to voicing rather than manner (Lisker, 1957b). The context was that of "rabid" ya "rapid." The results were: (1) that variation in the duration of intersyllabic silence was sufficient to cue the voicing distinction between the two words, and (2) that the location of the voicing boundary on the continuum of intersyllabic silence varied as a function of whether the stimuli were synthesized with or without a transition of the first formant at the end of the first syllable. Thus, cues with different acoustic properties were nevertheless found to be equivalent in phonetic perception: Just as stimuli characterized by the presence of a transition of the first formant and a relatively long silent interval were heard as "rapid," so also were stimuli characterized by the absence of a transition of the first formant and a shorter silent interval.

We ask now why silence should give rise to the same phonetic percept as the frequency modulation of the first-formant transition. As long as we think in terms of what we know, or can surmise, about auditory perception, the answer is elusive. Articulation, however, provides the tie that binds: these acoustically dissimilar events are both to be found among the many acoustic consequences of the gesture that converts "rabid" to "rapid." There are other, equally diverse acoustic consequences of the gesture, and these, too, according to the results of the early study and its current extensions (Lisker, 1977), have an equivalence in phonetic perception.

Since articulatory gestures commonly have multiple and diverse acoustic consequences, we should expect to find many cases of such perceptual equivalence among acoustically dissimilar cues. To be sure, there is no problem in finding such cases; they abound, and have been studied for all three phonetic dimensions: manner, voicing, and place. (For a review, see Liberman & Studdert-Kennedy, 1978.) In the third experiment of this section we examine one additional case. Taking advantage of the fact that the stop gesture that differentiates fricative from affricate in "ditch" ya "dish" generates changes in both the duration of the silent closure interval and changes in the onset and duration of the fricative noise, we examine the perceptual equivalence between silence, on the one hand, and, on the other, the rise time of the friction and its duration as well.

EXPERIMENT III

Our third experiment was designed to determine whether the perception of "split" could be induced by inserting silence between the fricative noise of [s] and the syllable "lit." Is silence, in this sense, a sufficient condition for the perception of stop manner, and, if so, over what range of durations is silence effective? The second question is interesting because we know that neither a very brief nor a very long closure is appropriate for stop manner. A too-brief closure would presumably indicate that the speaker had not closed his vocal tract long enough to have said "split." A too-long closure, on the other hand, would suggest that he had produced the "s," then waited a while, and finally said "lit." That being so, we should suppose that only a limited range of silent intervals would signal the production of stop manner.

Method

A male speaker's recording of the fricative noise of [s] and the syllable "lit" were digitized and stored in computer memory. (Both segments were produced in isolation.) Having listened carefully to these segments, we judged that the noise of the [s] did not end with a stop, nor did the "lit" begin with a stop. Using the editing facilities provided by the Haskins Laboratories PCM system, we then appended the "s-noise" to the "lit," separating these two segments by intervals of silence that ranged from 0 to 100 msec in steps of 15 msec, and from 100 to 650 msec in steps of 50 msec. Three tokens of each stimulus were generated. The resulting stimuli were randomized and recorded on audio tape with a three-sec interval between stimuli. The listeners were instructed to label the stimuli as "slit," "split," or "s" followed by "lit." (The last named category is not "slit," but rather "s" plus "lit," with a clearly perceptible period of silence in between.)

The subjects were 10 volunteers, all undergraduates at Lehman College who had not previously served in experiments on speech perception. They were tested in two groups (of five each) under conditions similar to those of Experiment I. To familiarize the listeners with the stimuli, we had them listen to the entire stimulus continuum before the test sequence began.

Results and Discussion

The effect of inserting intervals of silence between "s-noise" and [lit] is shown in Figure 10. There we see that at silent intervals of less than 60 msec listeners reported "slit," but at longer intervals--to about 450 msec--they reported "split." In this case, then, silence is a sufficient condition for stop manner. Notice, however, that at the longest silent interval the stop was not heard; rather, the subjects reported "s-silence-lit." Thus, neither the very brief nor the very long silent intervals produced a stop percept. This outcome accords well with our earlier supposition that only a limited range of silent intervals should signal stop manner.

EXPERIMENT IV

To this point we have investigated silence as a condition for the perception of stop manner. Now we turn to silence as a condition for affricate manner. To see why, consider that just as a speaker must close his

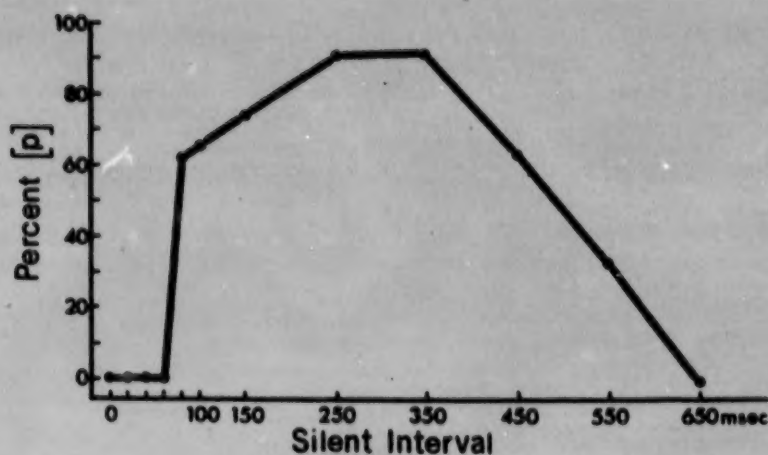


Figure 10. Silence as a sufficient condition of stop manner; identification of [p] in patterns composed of "s" followed by "lit."

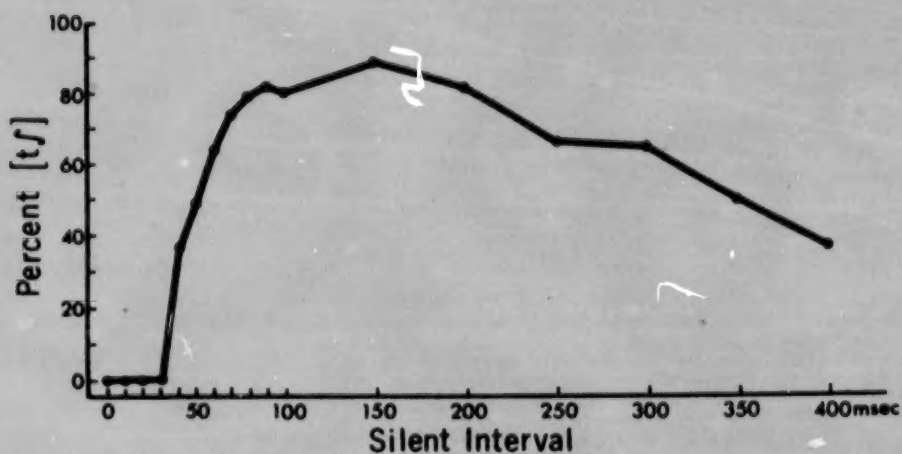


Figure 11. Silence as a sufficient condition for affricate manner; identification of [tʃ] in patterns composed of "please say" followed by "shop."

vocal tract to produce the stop that distinguishes, for example, [sta] from [sa], so also must he close his vocal tract to produce the stop-initiated fricative (that is, affricate) that distinguishes, for example, the phrase "say chop" from "say shop." There is evidence, moreover, that the silence associated with vocal-tract closure is a cue for the affricate-fricative contrast in intervocalic position. This evidence comes from early experiments with synthetic speech (Kuipers, 1955; Truby, 1955). The purpose of the experiment to be described here is to replicate and expand these early findings. Specifically, we aim to determine whether silence can be a sufficient condition for the fricative-affricate contrast in the naturally produced utterances "say shop" and "say chop."

Method

A male speaker's recording of "please say shop" was digitized and stored in computer memory. Using the editing facilities provided by the Haskins Laboratories PCM system, we removed the initial 15 msec of $\{$ -noise from "shop." The signal that remained still sounded to us like "shop."

We should note parenthetically that in situations of this kind, where there are presumably a number of different cues for the same distinction, it often happens that relatively extreme "settings" of one of the cues will cause the other cues to be "overridden" in perception. For example, in this case, we have reason to believe that the duration and onset of the frication noise, as well as silence, are cues to the affricate-fricative distinction (see Gerstman, 1957). Very long fricative noise, especially when combined with slow onset, may so bias perception toward the fricative that no amount of the silence cue can be effective.

To generate our experimental stimuli we inserted intervals of silence between the offset of "please say" and the onset of "shop." These intervals covered the range 0 to 400 msec. The steps were 10 msec each from 0 to 100 msec and 50 msec each from 100 to 400 msec. Four tokens of each stimulus were generated. The resulting stimuli were randomized and recorded on audio tape with a four-sec interval between stimuli.

The subjects were 10 volunteers, all undergraduates at Lehman College who had not previously participated in experiments on speech perception. They were tested en masse under listening conditions similar to those of Experiment I. The subjects were told they would hear either "please say shop" or "please say chop," and were instructed to write either "shop" or "chop" on their response sheets. To familiarize them with the experimental stimuli, we played twenty of the stimuli before the test sequence began.

Results and Discussion

The effect of varying the duration of the silent interval between "please say" and "shop" is shown in Figure 11. We see that "chop" responses begin to appear when the silent interval exceeds about 30 msec; by 70 msec they account for 75 percent of the responses. Thus, we conclude that silence can be a sufficient cue for distinguishing the affricate [tʃ] from the fricative [ʃ]. We should remark that, according to preliminary research we have done, the contrast between the voiced counterparts of those phones (that is, [dʒ] and

[ʒ]) can also be cued by silence.

Re-directing our attention to the data for the voiceless forms shown in Figure 11, we see that at the very long intervals of silence there is a tendency for our listeners' perceptions to revert to the fricative [ʃ]. This tendency is similar to what we saw in the case of silence as a cue for stop manner in the contrast "split" *ya* "slit" (cf. Figure 10), but it is not so marked. In that connection we note that the longest silent interval for the present experiment with "shop" and "chop" was 400 msec, whereas for the earlier experiment with "slit" and "split" it was 650 msec. When we examine the identification functions for "slit" *ya* "split," we see that at 400 msec our listeners' responses had only just begun to revert to "s-silence-lit." Presumably, then, in the present experiment, the "chop" responses would have reverted more nearly to "shop" had we carried the silent interval to greater lengths.

Having seen that the utterance "please say shop" can be converted into "please say chop" by appropriately increasing the silent interval between "say" and "shop," the question arises whether the utterance "please say chop" can be converted to "please say shop" by shortening the silence. The results from preliminary research suggest that this can be done, though just how convincingly depends upon the "intensity" of the affricate articulation in "chop" (Raphael & Dorman, 1977). Of course this is analogous to the results obtained in Experiments I and II, where too little silence caused stops not to be heard.

EXPERIMENTS Va AND Vb

Having found silence to be sufficient for the perception of affricate manner in syllable-initial position ("shop" *ya* "chop"), we now wish to determine whether it can be sufficient in syllable-final position, as in "dish" *ya* "ditch." We also wish in these experiments to examine the effects of two other cues for affricate manner--namely, the duration and rise-time of the fricative noise (see Gerstman, 1957)--and to study such relations as may exist between these two cues, on the one hand, and silence on the other.

Method

To provide a basis for the stimuli of Experiments Va and Vb, we twice recorded a male speaker saying "put it in the dish." These recordings were digitized and then stored in computer memory. To produce the experimental variation of primary interest we used the PCM editing system to introduce varying durations of silence between the end of voicing associated with the vowel [ɪ] and the beginning of the noise of [ʃ]. These durations ranged from 20 to 150 msec in steps of 10 msec. To enable us to study the effects of the silence cue in combination with the cues of duration and rise-time of the fricative noise, we introduced the silence cue into two series of stimuli. In Experiment Va we combined the silence cue with each of two durations of fricative noise, 320 msec and 160 msec, using for this purpose one of the two recordings referred to above. We produced the two durations of noise in the following way. For one, we used the noise of the original utterance, which was 320 msec in duration. To produce the other, which was 160 msec in duration, we removed 160 msec of noise from the center and then rejoined the

out ends. That operation obviously affects neither the onset nor offset characteristics of the noise.

In the other series we combined the silence cue with each of two different conditions of noise rise-time, using the second of the recordings referred to above. We produced the two rise-times in the following way. For one, we simply used the rise-time of the original utterance, which was 35 msec. For the other, we reduced the rise-time to an effective value of 0 msec by removing the first 30 msec of the noise. To compensate in the simplest possible way for the resulting reduction in overall duration of the noise, we added 30 msec of noise to the center. (Given that the rise-time was not instantaneous, this operation does not ensure that the durations of the stimuli with the two conditions of rise-time were psychologically equal. We should note, however, that they were more nearly so than they would have been if the 30-msec insertion had not been made.)

The subjects for Experiment Va were 10 undergraduate volunteers from Arizona State University who had not previously participated in research on speech perception. They were tested en masse in a large sound-attenuated room. The experimental stimuli were reproduced on a Magnecord 1032 tape recorder via a CEI 41-2 loudspeaker. The subjects for Experiment Vb were 12 undergraduate volunteers from Lehman College who had not previously participated in research on speech perception. They were tested in groups of four under the conditions described for Experiment I. The subjects in both experiments were given the same instructions. They were told that they would hear either "put it in the dish" or "put it in the ditch" and were instructed to write either "sh" or "ch" on their response sheets. In order to become familiar with the experimental stimuli, the subjects listened to twenty stimuli before starting the test sequence.

Results and Discussion

We see the results of Experiment Va in Figure 12. It is apparent that silence is sufficient in this case to cue the distinction between fricative and affricate postvocally. At the short intervals of silence the stimuli in both conditions of fricative-noise duration were heard as "dish," while at the longest intervals of silence they were heard as "ditch."

It is also apparent that there is a relation between the duration of silence and the duration of fricative noise. Thus, if we look at the silent interval necessary for 75 percent "ditch" responses, we see that it is approximately 89 msec when the noise is long (320 msec), but only 75 msec when the noise is short (160 msec). The difference in silent interval is significant ($T = 0, p < .005$). That is to say that 14 msec of silence (the difference between 89 msec and 75 msec) is equivalent in these phonetic perceptions to 160 msec of noise.

In Figure 13 we see the results of Experiment Vb. Since listeners report "dish" at the shortest intervals of silence and "ditch" at the longest intervals, we see, once again, that silence is sufficient to distinguish between fricative and affricate. Here, too, we see a relation between two acoustic cues to the same distinction: silence and rise-time of the fricative noise. The boundary between fricative and affricate is at about 57 msec of

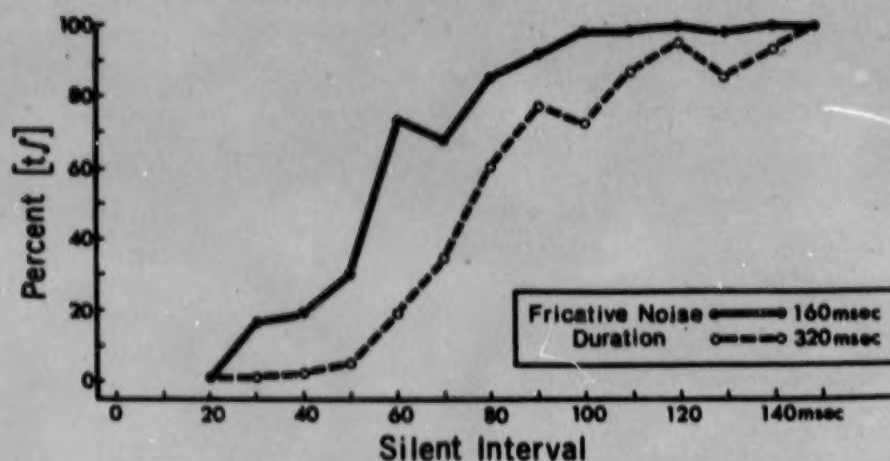


Figure 12. The relation between silence and sound; identification of [tʃ] for two conditions of fricative-noise duration.

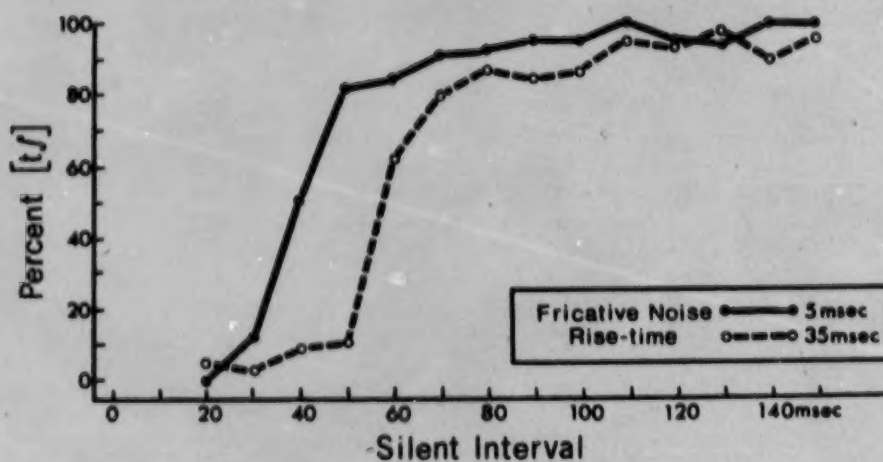


Figure 13. The relation between silence and sound; identification of [tʃ] for two conditions of fricative-noise rise-time.

silence when the rise-time is slow (35 msec), but at 37 msec when the rise-time is rapid (5 msec). This difference is significant ($T = 1$, $p < .005$).

We should note that relations of the kind described here can limit the effectiveness of silence as a cue. At one extreme we might have such a long duration of noise, and thus a strong bias toward a fricative, that no amount of silence would be sufficient to overcome it. At the other extreme there might be such a short duration and rise-time of the noise, and thus so strong a bias toward the affricate, that even durations of silence near 0 msec would not alter the perception of the affricate. This is consistent with the caveat we mentioned in the earlier discussion. It would apply also in the case of "slit" and "split" to the trading relation between temporal (silence) and spectral cues that have been reported by other investigators (Erickson, Fitch, Halwes, & Liberman, 1977; Liberman & Pisoni, 1977).

Returning now to the main findings of the experiment, we should note that the relations among the effects of the several cues are, in principle, like those that have been reported for numerous others (for a review, see Liberman & Studdert-Kennedy, 1978). In all cases, cues that are quite different from an acoustic point of view nevertheless give rise to the same phonetic percept. It is consistent with our hypothesis to suppose that the perceptual equivalence of these cues is due to the fact that they are the common products of the same linguistically significant gesture.

HOW THE EFFECTIVENESS OF SILENCE DEPENDS ON WHETHER IT COMES FROM ONE VOCAL TRACT OR TWO: AN ECOLOGICAL FACTOR IN PHONETIC PERCEPTION

Having suggested that silence is important in stop perception because it provides information about the behavior of a vocal tract, we should now ask: Whose vocal tract? We think it could hardly be that of the listener, nor of the speaker, nor of any particular person. Rather, it must be some more abstract conception of the behavior of vocal tracts in general. At all events, it is possible to find out; we need only take advantage of certain facts about the ecology of speech.

Consider two of the examples we developed in the earlier parts of this paper. First there was the case of [bab dɔ] and [bɔg dɔ], where it was found that a syllable-final stop was not perceived when there was an insufficiently long period of silence between the syllables. It was assumed that this was so because the relatively short silence informed the listener that the speaker must not have closed his vocal tract long enough to have produced a syllable-final stop. What one speaker cannot do, two speakers can: Given collaboration between two speakers, given the accidents of speech when several are talking, the utterance [bab dɔ], for example, can be produced with no silence at all between the syllables. Therefore, with two speakers (or more generally two sources of speech) the presence or absence of silence has no phonetic significance.

Similar considerations apply to our finding that the phrase "please say shop" was heard as "please say chop" when silence was inserted between "say" and "shop." By our account, the silence told the listener that the speaker had closed his vocal tract in a manner appropriate to the production of an affricate; hence, the perception of an affricate. Here, too, the presence or

absence of silence provides information only when there is but one speaker, for two can produce "please say" and "chop" with no silence at all between the words "say" and "chop."

Thus, silence does, or does not, provide useful phonetic information depending on whether (and how) the utterance was produced by one speaker or by two. The aim of the experiments to be reported here is to determine if listeners behave accordingly.

EXPERIMENT VI

The purpose of this experiment was to discover whether the effect of intersyllabic silence on the perception of syllable-final stops in the disyllables [bab da] and [bag da] is different when the syllables are produced by two speakers instead of one.

Method

Except for the introduction of a "different voice" condition, the procedures of this experiment were similar to those of Experiments IIa and IIb, where we were concerned with the effect of intersyllabic silence on the perception of syllable-final stops in [bab da] and [bag da]. First, we recorded a male saying [bab], [bag] and [da]. Those utterances were digitized and stored in computer memory. We then modified the [bab] and [bag] syllables by removing all but 15 msec of the voicing that followed the final formant transitions. To create the set of stimuli for the "same-talker" condition, we appended syllable [da] to [bab] and [bag] so as to create intersyllabic intervals of silence from 0 to 90 msec in steps of 10 msec. Three tokens of each stimulus were generated. The entire sequence was then randomized and recorded on audio tape with a three-sec interval between stimuli. To generate stimuli for the "different-talker" condition, we followed exactly the same procedure, but substituted a female voice saying [da]. Thus, we produced disyllables in which the first syllable ([bab] or [bag]) was in a male voice and the second syllable [da] in a female voice.

The subjects were 10 volunteers, all undergraduates at Lehman College who had previously participated in Experiment I. For the "same-talker" condition, the subjects were told that they would hear a male voice saying [bab da], [bag da] or [ba da]. For the different-talker condition, the subjects were told that they would hear a male voice saying [bab], [bag] or [ba] followed by a female voice saying [da]. In both conditions, the subjects were asked to identify, in writing, the final sound ([bab da], [bag da], or [ba da]) in each first (male-produced) syllable. The stimuli of the same- and different-talker conditions were presented in blocks. To control for practice effects, the order of the blocks was counterbalanced across the listeners. To familiarize the listeners with the stimuli, we presented twenty stimuli before each trial block.

Results and Discussion

The results for the same- and different-talker conditions are shown in Figure 14. Looking first at the same-talker condition, we see a result very similar to the one obtained in the analogous condition of one of our earlier

experiments (Experiment IIb): At short intervals of silence listeners did not hear syllable-final stops; these were heard with 75 percent accuracy only when the silent interval was about 45 msec in duration.

The result of the different-talker condition is quite different. Eight of the ten subjects identified syllable-final stops with near perfect accuracy at every interval of silence, including even the very shortest. For these subjects, it is as if their perceptual machinery "knew" that, with two speakers, intersyllabic silence conveys no useful phonetic information. The remaining two subjects behaved in the different-talker condition almost exactly as they had when there was but a single talker. We cannot be sure why. We may note, however, that a single syllable by each talker provides very little information about the identity of the talker. Conceivably, therefore, the fact that the two syllables were produced by different talkers did not properly "register" with these two subjects. In that connection, it is relevant that one of these two subjects did remark at the end of the experiment that she thought she had been listening to the same talker speaking on two different pitches. This suggests that the effect we obtained in the different-talker condition was not due solely to the acoustic differences between the voices as such, but rather to their role in informing the listeners that there were, indeed, two sources of speech.

EXPERIMENT VII

The purpose of this experiment was to determine if the effect of silence in converting "say shop" to "say chop" is different when the words on either side of the silence are produced by two talkers instead of one.

Method

The stimuli for this experiment were produced in the same manner as those of Experiment IV, except for the addition of a "different-voice" condition. First we digitized and stored in computer memory a male speaker's recording of "please say shop." To produce stimuli for the same-talker condition, we imposed intervals of silence between "please say" and "shop" in 10-msec steps over the range 0 to 100 msec. Three tokens of each stimulus were recorded. The entire sequence was then randomized and recorded with a three-sec interval between stimuli. To produce stimuli for the different-talker condition, we first digitized a female's recording of "please say shop." The phrase "please say" was excised from the recording and stored in computer memory. We then appended the male-produced "shop" to the female-produced "please say," leaving intervals of silence between "say" and "shop." These intervals ranged from 0 to 100 msec in steps of 10 msec. Three tokens of each stimulus were generated. The resulting stimuli were randomized and recorded on audio tape with a three-sec interval between stimuli.

The subjects were 10 volunteers, all undergraduates at Lehman College who had not previously participated in research on speech perception. For the "same-talker" condition, the subjects were told that they would hear a male voice saying either "please say shop" or "please say chop." For the different-talker condition, the subjects were told that they would hear a female voice saying "please say" and a male voice saying either "shop" or "chop." In both conditions the subjects were asked to write either "sh" (for "shop") or "ch"

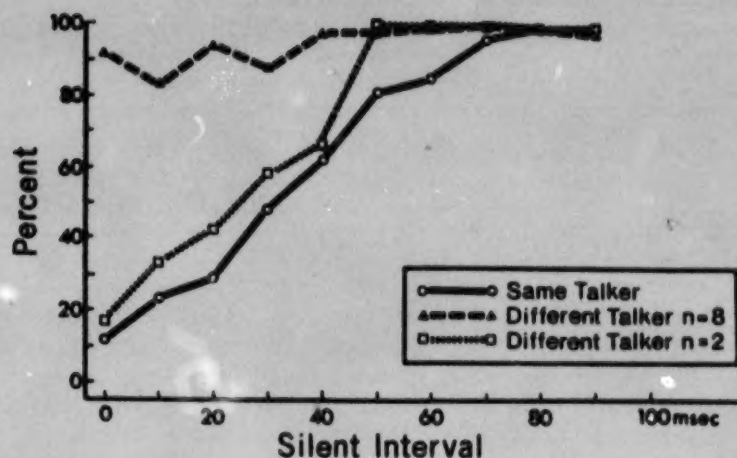


Figure 14. Silence as a condition for stop manner when it reflects the behavior of one vocal tract or two: identification of syllable-final stops in [bab d_a] - [bag d_a] in the same- and different-talker conditions.

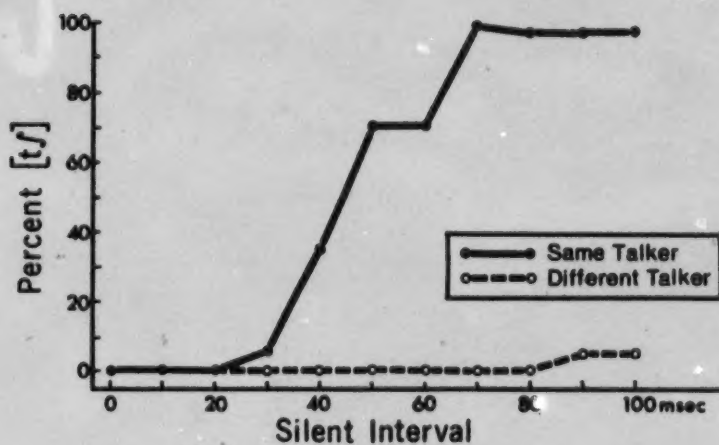


Figure 15. Silence as a condition for affricate manner when it reflects the behavior of one vocal tract or two: identification of [tʃ] in patterns composed of "please say" and "shop" in the same- and different-talker conditions.

(for "chop") on their response sheets. The subjects were tested in two groups of five under the listening conditions described in Experiment I. The stimuli of the same- and different-talker conditions were presented in blocks. The order of the blocks was counterbalanced across the two groups of subjects. To familiarize the subjects with the stimuli, we presented twenty stimuli before each trial block.

Results and Discussion

The results of Experiment VII are shown in Figure 15. One sees in the same-talker condition a result similar to that we obtained in the analogous condition of Experiment IV: The fricative in the word "shop" was heard as the affricate in the word "chop" when the silent interval between it and the immediately preceding word exceeded about 45 msec. In contrast, silence had no effect in the different-talker condition: Increases in the silent interval did not convert "shop" to "chop."

We should note that the utterance "please say shop" used in this experiment should have provided more information about the identity of the talker (or talkers) than did the two syllables of the previous experiment. This may account for the fact that, in this experiment, though not in the other, the effect of the same- versus different-talker conditions was found in every subject. Perhaps, however, the effect would not have been so large had we used other settings of the cues for the fricative-affricate distinction. Obviously, further research is necessary to determine the limits over which the effect obtains. We should also wonder about the effect in connection with the trading relations among the fricative-affricate cues that we observed in our earlier experiments. Having found, for example in Experiment 5, that duration of silence can be traded for friction duration, we might ask whether these cues also trade with the (perceived) magnitude of the difference between the voices.

We should emphasize that in both experiments the two talkers were male and female. Thus, the acoustic difference between the voices was relatively large. We are now conducting experiments contrived specifically for the purpose of determining whether the phenomenon here described depends critically on such an acoustic difference, or, alternatively, on an inference by the listener that he did or did not hear different voices. At this point, we believe it is the latter.⁵

GENERAL DISCUSSION

We should now assemble the results of our experiments in terms of their bearing on the three questions raised at the very beginning. As for the first question--Is silence a cue to stop manner?--the answer is quite straightforward and wholly in accord with the results of previous research: silence is a cue, necessary in some cases, sufficient in others. Thus, given spectral cues appropriate for a stop in absolute initial position (for example, [gg]), silence preceding those cues was found to be necessary if a stop was to be perceived as the second element of a fricative-stop-vowel syllable (for example, [ʃkɛ]). Similarly, in the case of stops in syllable-final position (for example, [bɛb]), silence following the spectral cues was necessary if they were to give rise to the perception of a stop when a second syllable was

added (for example, [bgb dɛ]). More interesting, perhaps, is the finding that even in the absence of sufficient spectral cues, silence did, in some circumstances, produce the perception of a stop or affricate. Thus, prefixing the noise of [s] to the syllable "lit" produced "split" when the correct amount of silence was interposed; inserting silence between the words "say" and "shop" converted them to "say chop."

The second question asked whether the effect of silence was exclusively auditory or also phonetic. If auditory, we should expect to find explanations in terms of masking or any one of a variety of interactions. If phonetic, we should assume that silence informs the listener that the speaker did or did not make the closure that is the distinguishing characteristic of the stops, and further that the listener is sensitive to that information, just as he would be if his perception of speech were constrained by knowledge of what a vocal tract must (or must not) do when it makes a linguistically significant gesture. This question is, by its nature, more problematic than the first one, and the answer is correspondingly harder to find. We believe, however, that the pattern of results obtained in the experiments reported here lends support to the assumption that the effect of silence is, to a significant extent, phonetic. Having presented these data at various places in this paper, we should collect them here.

First, we should consider again the basic fact that silence was an important cue, and then note how difficult it is, given our results, to account for that solely in auditory terms. Thus, we found that the transition cues for the stops were neither appreciably masked nor altered by interaction when, having been isolated from the speech patterns, they were heard as nonspeech chirps. It is also relevant, of course, that, under some conditions, silence was a sufficient cue. There were, in those cases, no other sufficient cues to be masked. It is also telling that silence was effective as a cue only over a limited range, just as should be expected given the assumption that it provides information about a stop closure that lasts for a limited amount of time. Further evidence for a link between perception and production is provided by those of the experiments that showed an equivalence in phonetic perception between duration of silence and duration of friction (or between duration of silence and the rise-time of the friction). That result, similar to the results of other investigators, seems easiest to interpret on the assumption that the acoustically different cues give rise to the same phonetic percept because they are normally the correlated (but distributed) acoustic consequences of the same gesture.

Having said that the data of our experiments (and those of others) imply that perception of the silence cue is constrained as if by knowledge of what vocal tracts can do, we should offer a few parenthetical comments about what the data do not imply. First, they most certainly do not imply that a listener can hear only what a vocal tract can do. Indeed, it is for that reason that we have so often added the qualification "when the vocal tract makes linguistically significant gestures." For we know that synthetic speech can be readily perceived (as speech), though it departs, sometimes appreciably, from those acoustic patterns that real vocal tracts can produce. Thus, synthetic patterns sometimes contain only two formants, and the transitions are sometimes made to change direction instantaneously. But such departures, we should note, are not linguistically relevant. Languages cannot enforce a

distinction between phones made with two formants and those made with the greater number of formants that real vocal tracts produce, nor can they contrast instantaneous changes in formant slope with those more gradual changes that must characterize the behavior of such real masses as the tongue. In cases like these, an experimenter can take all manner of liberties with the stimulus patterns without destroying or even distorting phonetic perception, provided he manages to include the acoustic information that enables the listener to hear the stimuli as speech. All this is to say that if the speech perceiving mechanism is "tuned" to a vocal tract, as implied, then such "tuning" must hold only for those maneuvers that have linguistic significance.

Second, the assumption of a link between perception and production is not meant to imply anything about the nature of the mechanism that mediates the link or about the relative contributions of nature and nurture to its formation. In regard to the nature of the mechanism, there are aspects of our results (and those of others) that speak against at least one very simple possibility: feature detectors that have evolved in such a way as to be "tuned" to respond to fixed acoustic consequences of articulatory gestures and to be "sprung" when those consequences are present in the signal. In that connection, we note, first, that the relations among cues that we have found suggest that the setting of one detector (for example, the silence detector) must be, in effect, variable and conditioned by the "value" of the other cues (for example, duration of the noise). We should then note that, according to the results of the experiment on identification of syllable-final stops, a detector for the syllable-final transition cues could not respond directly upon sensing these cues, but would, instead, have to wait until it had information about the next syllable. At the least, it would have to know about that next syllable how far removed in time it was from the syllable containing the target phone and what kinds of phones it comprised. The consequence for a detector model is that it loses much of the appeal that it would otherwise have by virtue of its simplicity.

As for questions about the contributions of nature and nurture to the assumed link between perception and production, we should emphasize that such questions stand apart from those that pertain to the existence of such a link. Our experiments bear only on the latter.

We turn finally to the third question: Whose vocal tract is perception linked to? Given the results of the experiments with same and different talkers, we should suppose that the answer is quite clear: The relevant vocal tract is not that of the listener nor is it that of the speaker; it is rather some very abstract conception of vocal tracts in general. However, those same results add support to the view that a link to some vocal tract, however abstract, does figure in the perception of speech.

REFERENCE NOTE

1. Bailey, P., Summerfield, Q., & Dorman, M. Friction duration and friction offset as cues to stop manner in fricative-stop-vowel sequences. In preparation.

REFERENCES

- Abbs, M. A study of cues for the identification of voiced stop consonants in intervocalic contexts. Unpublished doctoral dissertation, University of Wisconsin, 1971.
- Bastian, J. Silent intervals as closure cues in the perception of stops. Haskins Laboratories, Speech Research and Instrumentation, 1962, 9, Appendix F.
- Bastian, J., Eimas, P., & Liberman, A. Identification and discrimination of a phonemic contrast induced by silent interval. Journal of the Acoustical Society of America, 1961, 33(A), 842.
- Darwin, C. J. Ear differences in the recall of fricatives and vowels. Quarterly Journal of Experimental Psychology, 1971, 23, 46-62.
- Darwin, C. J., & Bethell-Fox, C. Pitch continuity and speech source attribution. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 665-672.
- Erickson, D., Fitch, H., Halwes, T., & Liberman, A. Trading relation in perception between silence and spectrum. Journal of the Acoustical Society of America, 1977, 61(A), 546.
- Fujisaki, H., Nakamuro, K., & Imoto, T. Auditory perception of duration of speech and non-speech stimuli. In G. Fant & M. A. A. Tatham, (Eds.), Auditory analysis and the perception of speech. London: Academic Press, 1975.
- Ganong, W. An experiment on "phonetic adaptation." MIT Research Laboratory of Electronics, Progress Report, 1975, 116, 206-210.
- Gerstman, L. J. Perceptual dimension for the friction portions of certain speech sounds. Unpublished doctoral dissertation, New York University, 1957.
- Harris, K. S. Cues for the discrimination of American English fricatives in spoken syllables. Language and Speech, 1958, 1, 1-7.
- Kuipers, A. Affricates in intervocalic position. Haskins Laboratories Quarterly Progress Report, 1955 15, Appendix 6.
- Liberman, A. M., & Pisoni, D. B. Evidence for a special speech-processing subsystem in the human. In T. H. Bullock (Ed.), Recognition of complex acoustic signals (Life-Sciences Research Report 5). Berlin: Dahlem Konferenzen, 1977.
- Liberman, A. M., & Studdert-Kennedy, M. Phonetic perception. In R. Held, H. Leibowitz, & H. L. Teuber (Eds.), Handbook of sensory physiology, (Vol. VII), Perception. Heidelberg: Springer-Verlag, 1978.
- Lisker, L. Closure duration and the voiced-voiceless distinction in English. Language, 1957, 33, 42-49. (a)
- Lisker, L. Closure duration, first-formant transitions and the voiced-voiceless contrast of intervocalic stops. Haskins Laboratories Quarterly Progress Report, 1957, 23, Appendix 1. (b)
- Lisker, L. Closure hiatus: cue to voicing, manner and place of consonant occlusion. Journal of the Acoustical Society of America, 1977, 61(A), S-48.
- Pickett, J. M., & Decker, L. Time factors in perception of a double consonant. Language and Speech, 1960, 3, 11-17.
- Port, R. The influence of speaking tempo on the duration of stressed vowel and medial stop in English trochee words. Unpublished doctoral dissertation, University of Connecticut, 1976.
- Raphael, L. J., & Dorman, M. F. Perceptual equivalence of cues for the

- fricative-affricate contrast. Journal of the Acoustical Society of America, 1977, 61(A), S-45.
- Repp, B. Perception of implosive transitions in VCV utterances. Haskins Laboratories Status Report on Speech Research, 1976, SR-48, 209-234.
- Repp, B. Perceptual integration and selective attention in speech perception: Further experiments on intervocalic stop consonants. Haskins Laboratories Status Report on Speech Research, 1977, SR-49, 37-70.
- Rudnicki, A., & Cole, R. Vowel identification and subsequent context. Journal of the Acoustical Society of America, 1977, 61(A), S39.
- Summerfield, A. Q., & Bailey, P. On the dissociation of spectral and temporal cues for stop consonant manner. Journal of the Acoustical Society of America, 1977 61(A), S-46.
- Truby, H. Affricates. Haskins Laboratories Status Report on Speech Research, 1955 11, 7-8.

FOOTNOTES

¹When native speakers of English produce [ʃp̥] and [ʃk̥], [p] and [k] are realized as voiceless inaspirates. It is for this reason that, when the fricative noise is removed from [ʃp̥] and [ʃk̥], listeners hear the stops that remain as voiced. In our experiment, it was necessary, therefore, to record [b̥] and [g̥] (rather than [p̥] and [k̥]), so that, when the fricative noise and vocalic segment were combined, the listeners would hear a normal sounding [ʃp] and [ʃk].

²The term "geminate" is ordinarily used to refer to the doubling of a consonant within a word. Such doubling as we find in English occurs only across word boundaries. We nevertheless use the term here, though our subjects were native speakers of English and were accustomed to consonant doubling only at word boundaries.

³Since writing this paper, a somewhat similar result by Rudnicki and Cole (1977) has come to our attention. Having recorded [ba ga], they found: (1) that after removing the [ga] their listeners heard [bag]; (2) that after replacing the [ga] with [da] placed close in time to the first syllable, listeners heard [bai da], and (3) that when the second syllable was separated from the first syllable by a sufficient interval of silence, listeners heard [bag da]. This result is of particular interest from our point of view because, in the condition when the second syllable [da] was close to [ba] and the subjects heard [bai da], it is clear that the transition cues at the end of the first syllable were not being (backward) masked by the second syllable; they were being perceived, but as a glide to [i] rather than as a stop. That result is similar to the finding of Liberman and Pisoni (1977), referred to earlier in this paper, that ʃ-noise placed close to [g̥] causes listeners to perceive [ʃj̥].

⁴We have not commented on the difference between the identification functions for [b] and [g] because we have found that difference to change, even to be reversed, depending on the surrounding vocalic environment. We emphasize the geminate vs nongeminate contrast because it remains more nearly stable across vowel environments.

⁵Using stimulus patterns and procedures very different from ours, Darwin and Bethell-Fox (1977) have, nevertheless, obtained results that are quite compatible. After synthesizing a pattern that was heard as an uninterrupted sequence of semivowels and vowels, they found that introducing changes in fundamental frequency at appropriate places in the pattern (without changing formant frequencies) caused the semivowels to be heard as stops. Their interpretation was that the rapid shifts in fundamental frequency caused the sequence to "stream," thus permitting the listener to hear two voices; that, in turn, provided the silence necessary to convert semivowel to stop.

PHONOLOGICAL CODING IN BEGINNING READING

Carol A. Fowler*

Abstract. Speech coding may contribute to the skilled reading process in at least two ways. Phonological short-term memory may facilitate comprehension of text, and the phonological form of a written word may serve as the word's lexical address. Research concerning correlates of beginning reading suggests that speech coding serves similar roles for the beginning reader. Good and poor beginning readers, and also, less and more experienced readers are distinguished on measures of linguistic awareness and on several other indicants of facility with speech coding.

INTRODUCTION

A word in a spoken language has two essential properties. It has a meaning or meanings and it has a phonological form. Neither a meaningless label nor an unencoded meaning can be a word of any language.

As many investigators have pointed out, written languages are parasitic on spoken ones. Thus, the patterning of symbols in written text makes reference to some corresponding patterning in a spoken language. In alphabetic writing systems, the primary correspondence is with the (meaningless) sound elements of the spoken language.

Since words of a language have essential phonological as well as semantic properties, when a reader recognizes a written symbol or symbol string as a referent of a word in his language he achieves access both to the semantic and to the phonological properties constituting the word. Consequently, it may be useless to debate whether or not phonological information is accessed during reading (i.e., whether reading can be a "purely visual" process).

It may be useful, though, to ask what role phonological information serves in reading. For readers of an alphabetic orthography, at least two roles are possible. Of them, one is probably essential, while the other may be optional--at least for skilled readers.

The essential function of phonological information is to provide a convenient form for the short-term storage of textual material while it is

*Also Dartmouth College, Hanover, New Hampshire.

Acknowledgment: This work was supported by NICHD Grant HD-01994 to Haskins Laboratories.

being read. In a grammar that allows embedding, close grammatical relationships may exist between words that are quite far apart in a sentence (e.g., "The woman who lives next door writes a column in the local newspaper"). On these grounds it has been argued that substantial portions of a sentence must be held in storage until a whole syntactic chunk is available for semantic processing. The memory system that is believed to have the requisite capacity and longevity for this purpose is phonological short-term memory.

Similarly, writers as well as talkers use pronouns in the place of common or proper nouns, but only when they can assume that the reader or listener is currently "thinking about" the pronoun's referent (e.g., Chafe, 1974). Thus, the sentence pair in 1 is natural, while that in 2 is not:

1. John broke the expensive vase. He tripped over it on his way out the door.
2. John broke the expensive vase. John tripped over the expensive vase on John's way out the door.

But pronominalization is only feasible if the reader or listener keeps old information (i.e., the information given in the first sentence of each pair) "in mind" as he processes new information. Again, the likely means of keeping old textual information available is phonological short-term memory.

These considerations are supported by data from several sources. For example, skilled readers of English persist in phonological access during reading of text even when doing so impairs performance on a second task ("g-cancellation," Corcoran, 1966). Moreover, preventing phonological access during reading impairs memory for meaning (Levy, 1975). Finally, skilled readers of a logographic orthography (Chinese) code a written text phonologically when their task is to read for meaning (Tzeng, Hung, & Wang, 1977). These studies suggest that eventual access to phonological information (before, or at the time of lexical access) may be an invariant aspect of skilled reading across orthographies and across reading tasks that require short-term memory.

Beyond providing a convenient temporary storage medium, however, phonological information may be involved in the reading process in a second way as well. Readers of an alphabetic orthography may gain access to the lexicon by applying spelling-to-sound rules to a written word in order to extract its corresponding phonological form. The phonological information may then serve as it does in listening as the "address" for the appropriate lexical entry. Baron and Strawson (1976) provide evidence that some skilled readers tend to use this means of word identification. Of course, the skilled reader need not access the lexicon in this way. Baron and Strawson (1976) also provide evidence that other readers habitually access the lexicon based on a word's orthographic form.

THE BEGINNING READER: THE POSSIBILITY OF FLEXIBLE PROCESSING STRATEGIES

But what of the beginning reader of an alphabetic orthography? What role does phonological coding serve in the beginner's efforts to read, and what role should it serve? Certainly, the requirements of short-term memory are as

critical to him as they are to the skilled reader. But what of its role in reading isolated words? Should the child exploit the sound-based patterning of the orthography in his reading of single-words, or should he bypass it as some skilled readers tend to do, and as readers of logographic writing systems apparently must do?

A moment's consideration suggests that he should be capable of doing both. Each processing style has its special advantages and disadvantages that may best suit it for different kinds of words or for different situations in which words are to be identified.

Consider first the strategy in which lexical access is based on a word's holistic optical form, and thus in which the sound-based patterning of the orthography is irrelevant to the reading task. Some words, namely those that do not conform to English spelling-to-sound rules, must be read in this way. An advantage of this strategy for the child may be that his reading can be more fluent than it is when he stops to sound out each word. However, at the very earliest stages of learning to read, this strategy places an enormous burden on the child's ability to memorize word shapes by rote, and on his ability to make intelligent guesses based on context when he sees an unfamiliar word.

For its part, the second strategy--of accessing the lexicon by way of the phonological form of a word--also has important advantages and disadvantages. Its main advantage is that it exploits the (mostly) ruleful relationship between orthography and sound. Thereby it enables the child to read most of the words that he knows by sound but not yet by sight.

As important as this advantage is, it is countered somewhat by two apparent disadvantages. One is, that for an unpracticed reader, application of sound-spelling rules is time-consuming. Thus, reading may not be fluent, and the child may have difficulty remembering words that he has already read as he is confronted with the "interfering" task of word decoding. The second disadvantage is that this strategy requires what Mattingly (1972) has called "linguistic awareness," and similarly, it requires that the child understand the relationship between aspects of the language and the orthography. Thus, it requires that the reader be aware that words, both written and spoken, have an internal structure, and that the internal structure of the one refers to (provides information about) the internal structure of the other. Recent research makes it quite clear that information about the sound structure of a spoken word or syllable is not readily available to the nonreader's awareness (Liberman, Shankweiler, Fischer, & Carter, 1974). Rather, the child has to learn explicitly what he already knows tacitly--namely that words are sequences of phonological segments. This is the problem of linguistic awareness. In addition, there is the related problem of understanding the relationship between analogous characteristics of written and spoken words. Even seemingly obvious relationships are not obvious to the young child. He may not know, for instance, that orthographic length is correlated with spoken duration (Rozin, Bressman, & Taft, 1974).

The foregoing considerations of the advantages and disadvantages of the two reading strategies suggest that an optimal approach to reading for a child is one in which he uses both means of lexical access--but in either case,

holds the outcome in phonological short-term storage. If he can recognize a word on sight, then that may be the most efficient means of access to the lexicon. Failing that, however, it is important that he have a reliable way to identify a word. The most reliable way, given the nature of the orthography, is the strategy of exploiting the ruleful relationship between atomic units of the orthography and those of the language.

Research assessing the processing styles of beginning readers

My colleagues and I have been concerned with examining the role of phonological coding in beginning reading, both in respect to its role in the short-term storage of verbal information (Liberman, I. Y., Shankweiler, Liberman, A. M., Fowler, & Fischer, 1977) including text (Fowler, Note 1; Mann, Liberman, Shankweiler, & Katz, Note 2) and in respect to its role in lexical access (Fowler, Liberman, & Shankweiler, 1977; Mark, Shankweiler, Liberman, & Fowler, 1977). We have tried to assess the importance of this linguistic aspect of reading for the beginner by comparing the effectiveness with which he accesses and uses phonological information to his degree of reading skill. We have devoted less effort to the visual component of the reading process primarily because the available evidence suggests that it is not a problematic aspect of learning to read, even among poor readers (e.g., Liberman & Shankweiler, in press; Vellutino, 1977). Three of the areas in which my colleagues and I have studied phonological processing in beginning reading are briefly summarized below.

Phonological coding in reading text

In any reading task that involves short-term memory—either explicitly or by implication in requiring comprehension of text—we should find that the beginning reader encodes the textual material phonologically. Moreover, given that good and poor readers do not differ strikingly on nonlinguistic aspects of the reading task, we should expect to find the differences among them to appear in the extent to which they make efficient use of the phonological representation in their reading of text.

Two results obtained by our research group bear out these predictions. In one experiment (Fowler, Note 1), second-grade good and poor readers were given two tasks to be performed concurrently. They were asked to read a short passage for comprehension and, at the same time, to cancel out any letter e that they saw while reading. The task was modeled after Corcoran's original experiment designed to assess phonological coding by adult readers. Corcoran found that silent e's were missed more often than nonsilent e's and suggested that this difference could arise only if the subjects were coding the written words into some sound-based form. Our study replicated Corcoran's in showing a higher proportion of silent e's being missed by both groups of readers than nonsilent e's. In our study, although good readers tended to show a larger silent e effect than poor readers, the difference did not approach significance.

However, in a study of immediate recall of sentences, Mann, Liberman, Shankweiler, and Katz (Note 2) did obtain the expected difference between good and poor readers. In this study, good readers were substantially more impaired than were poor readers by phonological confusability among the

component words of a sentence.

Linguistic awareness and beginning reading skill

Beyond the role of the phonology in comprehension and storage of text, we have suggested that phonological information may also be invoked by application of spelling-to-sound rules when individual words are read. This use of the sound-patterning of the language, as noted above, demands "linguistic awareness" on the part of the reader. Therefore, we would expect to find a relationship between a child's degree of linguistic awareness and his ability to read isolated words. Several studies have obtained a significant correlation between a subject's performance on Liberman's phoneme segmentation task (Liberman et al., 1974), designed to measure linguistic awareness, and performance on the Wide Range Achievement Test, which assesses skill in reading isolated words (Helfgott, 1976; Zifoak, 1977). Thus, subjects who perform more poorly when asked to indicate the number of phonemic segments in a word by tapping once per segment, also rank lowest on a test of isolated-word reading.

Phonological coding in the reading of isolated words by good and poor readers

If more and less skilled readers are distinguished either in the extent to which they use the phonological coding strategy of lexical access, or in the success with which they use it, the coding component in isolated-word reading might be expected to be more salient among good than among poor readers, and more salient among experienced than among less experienced readers.

A difference between good and poor readers was found in an experiment by Mark, Shankweiler, Liberman, and Fowler (1977). In that study, second-grade good and poor readers were given a list of words to read aloud. Following that, unexpectedly, they were given a recognition task including the words that they had just read and a set of rhyming and nonrhyming foils. Good, but not poor, readers made significantly more false positive responses to rhyming foils than to nonrhyming foils.

An investigation of the development of phonological coding skills provides compatible data (Fowler, Shankweiler, & Liberman, 1978) in showing a relationship between skill in accessing the phonological form of a letter string and reading experience. This study showed an increase in tendency to apply spelling-to-sound rules appropriately to nonsense letter-strings with increasing reading experience among second-, third-, and fourth-grade children.

Summary

The literature on skilled reading suggests two ways in which speech coding contributes to the skilled reading process. One is that phonological short-term memory facilitates the comprehension of text, and the other, that the phonological form of a written word may serve as the word's lexical address. Both of these services are at least as critical to the beginner as they are to the skilled reader. Our research concerning the correlates of beginning reading attests to this in showing that good and poor beginning

readers, and likewise, less and more experienced readers are distinguished on measures of linguistic awareness and on several other indicants of facility with speech coding.

REFERENCE NOTES

1. Fowler, C. A. Speech coding during reading of text by good and poor beginning readers. In preparation.
2. Mann, V., Liberman, I. Y., Shankweiler, D. P., & Katz, L. Children's memory for sentences and word strings in relation to reading ability. In preparation.

REFERENCES

- Baron, J., & Strawson, C. Use of orthographic and word-specific knowledge in reading words aloud. Journal of Experimental Psychology: Human Perception and Performance, 1976, 2, 386-393.
- Chafe, W. Language and consciousness. Language, 1974, 50, 111-133.
- Corcoran, D. W. J. An acoustic factor in letter cancellation. Nature, 1966, 210, 658.
- Fowler, C. A., Liberman, I. Y., & Shankweiler, D. P. On interpreting the error pattern in beginning reading. Language and Speech, 1977, 20, 162-173.
- Fowler, C. A., Shankweiler, D. P., & Liberman, I. Y. Apprehending spelling patterns for vowels: A developmental study. Haskins Laboratories Status Report on Speech Research, 1978, SR-57, 231-240.
- Helfgott, J. Phoneme segmentation and blending skills of kindergarten children: Implications for beginning reading acquisition. Contemporary Educational Psychology, 1976, 1, 157-169.
- Levy, B. A. Vocalization and suppression effects in sentence memory. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 304-316.
- Liberman, I. Y., & Shankweiler, D. P. Speech, the alphabet and learning to read. In L. Resnick & P. Weaver (Eds.), Theory and practice of early reading. Hillsdale, N. J.: Erlbaum, in press.
- Liberman, I. Y., Shankweiler, D. P., Fischer, F. W., & Carter, B. Reading and the awareness of linguistic segments. Journal of Experimental Child Psychology, 1974, 18, 201-212.
- Liberman, I. Y., Shankweiler, D. P., Liberman, A. M., Fowler, C. A., & Fischer, F. W. Phonetic segmentation and recoding in the beginning reader. In A. S. Reber, & D. Scarborough (Eds.), Toward a psychology of reading. Hillsdale, N. J.: Erlbaum, 1977.
- Mark, L. S., Shankweiler, D. P., Liberman, I. Y., & Fowler, C. A. Phonetic recoding and reading difficulty in beginning readers. Memory & Cognition, 1977, 5, 623-629.
- Mattingly, I. G. Reading, the linguistic process and linguistic awareness. In J. F. Kavanagh & I. G. Mattingly (Eds.), Language by ear and by eye: The relationships between speech and reading. Cambridge, Mass.: MIT Press, 1972.
- Rozin, P., Bressman, B., & Taft, M. Do children understand the basic relationship between speech and writing? Journal of Reading Behavior, 1974, 6, 327-334.

- Tzeng, O., Hung, D. L., & Wang, W. S.-Y. Speech recoding in reading Chinese characters. Journal of Experimental Psychology: Human Learning and Memory, 1977, 3, 621-630.
- Vellutino, F. Alternative conceptualizations of dyslexia: Evidence in support of a verbal-deficit hypothesis. Harvard Educational Review, 1977, 47, 334-354.
- Zifoak, M. Phonological awareness and reading acquisition in first-grade children. Unpublished doctoral dissertation, University of Connecticut, 1977.

LARYNGEAL ADJUSTMENTS DURING JAPANESE FRICATIVE AND DEVOICED VOWEL PRODUCTION*

Hirohide Yoshioka†

Abstract. The aim of the present paper is to clarify the role of laryngeal adjustments for phonetic variations of voicing in Japanese--vowel devoicing and intervocalic /h/ voicing--by use of electromyography (EMG) and fiberoptics. The results indicate that the phenomenon of vowel devoicing is accompanied by EMG activity patterns of the posterior cricoarytenoid and interarytenoid different from those for fully voiced vowels, causing the glottis to remain open. In contrast, the voicing of /h/ is quite different, in that it occurs while the glottis remains as wide as it does for voiceless /h/ or /s/ with comparable EMG patterns of those muscles, despite the presence of vocal fold vibration. Therefore, it may be that this latter phenomenon is chiefly dependent on some other condition at the level of the glottis. The paper also deals with some critical cases where either of these allophonic variations of voicing can occur, such as the /ih/ sequence in meaningful words like /si^hhee/ and /sihee/. The EMG data suggest that the lesser frequency of vowel devoicing for the accented nuclear vowel, /i^h/ in the former word for example, might be attributed to the rapid and high activity of the interarytenoid for this particular vowel, causing definite closure of the glottis and consequently allowing the excited vibrations to continue during the following /h/ segment in spite of the widely separated glottis.

INTRODUCTION

It has been well established that the larynx plays an important role in accomplishing voicing distinctions. The approximation of the vocal folds, in particular, is considered one of the crucial conditions for presence or absence of vocal fold vibration. Many studies using transillumination and fiberoptic techniques have confirmed that the precise timing control of this glottal opening and closing gesture in reference to the supraglottal articulatory movements is critically linked not only to the manifestation of voicing but aspiration as well. The EMG work has further confirmed that the degree and timing of the glottal aperture is controlled mainly by the activity patterns of the abductor and adductor muscle groups of the larynx. Therefore, voicing

*A version of the paper was presented at the Meeting of the American Speech and Hearing Association in San Francisco, November 1978.

†Also University of Tokyo, Japan.

Acknowledgment: This work was supported in part by NINCDS Grant NS-13870 and BRS Grant RR-05596 to Haskins Laboratories. I would like to thank Dr. Seiji Niimi for his participation as the subject. I am also grateful to Drs. Hajime Hirose and Anders LQqvist for their comments on the paper.

and/or aspiration reveal themselves as different activity patterns of the intrinsic laryngeal muscles at the EMG level (Lisker, Abramson, Cooper, & Schvey, 1969; Sawashima, 1970; Hircse & Gay, 1972; Sawashima & Miyazaki, 1973; Kagaya, 1974; Fischer-Jørgensen & Hirose, 1974; Dixit, 1975; Kagaya & Hirose, 1975; Iwata & Hirose, 1976; Hirose, 1976; Benguerel, Hirose, Sawashima, & Ushijima, 1978).

Besides phonemic distinctions, for instance in the Tokyo dialect of Japanese, there are a few explicit voicing variations from the phonetic viewpoint (e.g., Hattori, 1951; Han, 1962a; Fujimura, 1971; Sawashima, 1973). One is that the high vowels /i/ and /u/ between voiceless consonants are often devoiced. The other is that the fricative /h/ in an intervocalic position is frequently accompanied by vocal fold vibration. The present experiment was conducted to clarify the underlying physiological differences in laryngeal adjustment for these "nondistinctive voicing variations" in Japanese, by applying the EMG techniques to the intrinsic laryngeal muscles and the fiberoptic observation method to the upper view of the glottis. The author believes that this kind of study, focused on redundant articulatory gestures in terms of their relevance to phonemic distinctions, could provide some additional insight into the veiled portion of the biomechanism behind phonologically significant features.

METHOD AND PROCEDURE

The EMG data were obtained by use of the hooked-wire electrode techniques. The electrodes, consisting of a pair of platinum-tungsten alloy wires (0.002 in. in diameter with isonel coating), were inserted perorally into the posterior cricoarytenoid (PCA) and the interarytenoid (INT) under indirect laryngoscopy with the aid of a specially designed curved probe. For placement into the cricothyroid (CT), a percutaneous approach was adopted, using a hypodermic needle (26 gauge and 1 1/2 in. in length) as a guide (Hirano & Ohala, 1969; Hirose, Gay, Strome, & Sawashima, 1971; Hirose, 1971a).

The interference patterns of EMG signals were recorded on an FM multi-channel data recorder together with the acoustic signal. The action potentials were fed into a digital computer system and sampled at a rate of 200/sec, after being rectified and integrated over a 5 msec time window for further processing, to obtain the appropriate muscle activity patterns for single tokens and/or averaged ones (Kewley-Port, 1973, 1974, 1977).

For the fiberoptic data, the glottal view through the laryngeal fibroscope (4.5 mm in outer diameter) was photographed with a cine camera at a rate of 60 frames/sec simultaneously with the EMG and speech signals for some tokens of each utterance type. In each frame the distance between the vocal processes, an indicator of the glottal width, was measured (Sawashima & Hirose, 1968; Sawashima, 1977).

A native adult male speaker of the Tokyo dialect served as the subject. Among the possible phoneme sequences composed of /C₁iC₂ee/, /C₁eeC₂i/ and /C₁eeC₂ee/ (C₁, C₂ = h, s), meaningful words only were selected as test utterances. As is shown in the left portion of Table 1, some of these phoneme sequences may form two different words depending on whether the accent kernel is present or absent. The position of each accent kernel is indicated by the

diacritic symbol "̣". The subject was asked to pronounce each test word 28 times in a frame sentence, "sorewa ____ desu", "we call it ____", in random order. No particular mention was made of voicing variation. The vocal intensity, the pitch, and the speaking rate were also left to his discretion.

Table 1

List of the test words and occurrences of voicing variation.

TEST WORD	VOICED	VOICELESS
/hihee/	6	22
/hisee/	4	24
/sihee/	11	17
/sịhee/	27	1
/sisee/	5	23
/sịsee/	20	8
/heesi/	-	-
/hẹesi/	-	-
/seesi/	-	-
/sẹesi/	-	-
/heehee/	28	0
/heese/	-	-
/seehee/	28	0
/seese/	-	-

RESULTS

There was variation in pronunciation among the tokens of each utterance type, in that the vowel /i/ became devoiced or not and/or the intervocalic /h/ remained voiceless or not. These variations in voicing were detectable in sound spectrograms, in audio waveforms, and by the judgment of a phonetically trained listener.

The right portion of Table 1 shows a number of the tokens classified into the same subset group for some utterance types, with regard to the phonetic variations. It reveals the validity of previous impressions about the occurrence of the devoicing, that the nuclear vowel /i/ tends to be less frequently devoiced in an accented mora and vice versa. Productions are, however, variable. For example, the 8 tokens for the utterance type /sịsee/ were produced with devoiced /i/, and the 5 tokens for /sisee/ were fully voiced. On the other hand, intervocalic /h/'s with /e/'s on both sides were uttered with vocal fold vibration without exception. More details connected with the corresponding EMG data will be presented below.

Figure 1 contains the first 8 tokens of the 28 productions for the utterance type /hisee/. The left portion indicates the binary judgments in voicing variation. Tokens number 1 and 3 among this sample were judged as voiced, and all others as devoiced by aid of the above mentioned method. For

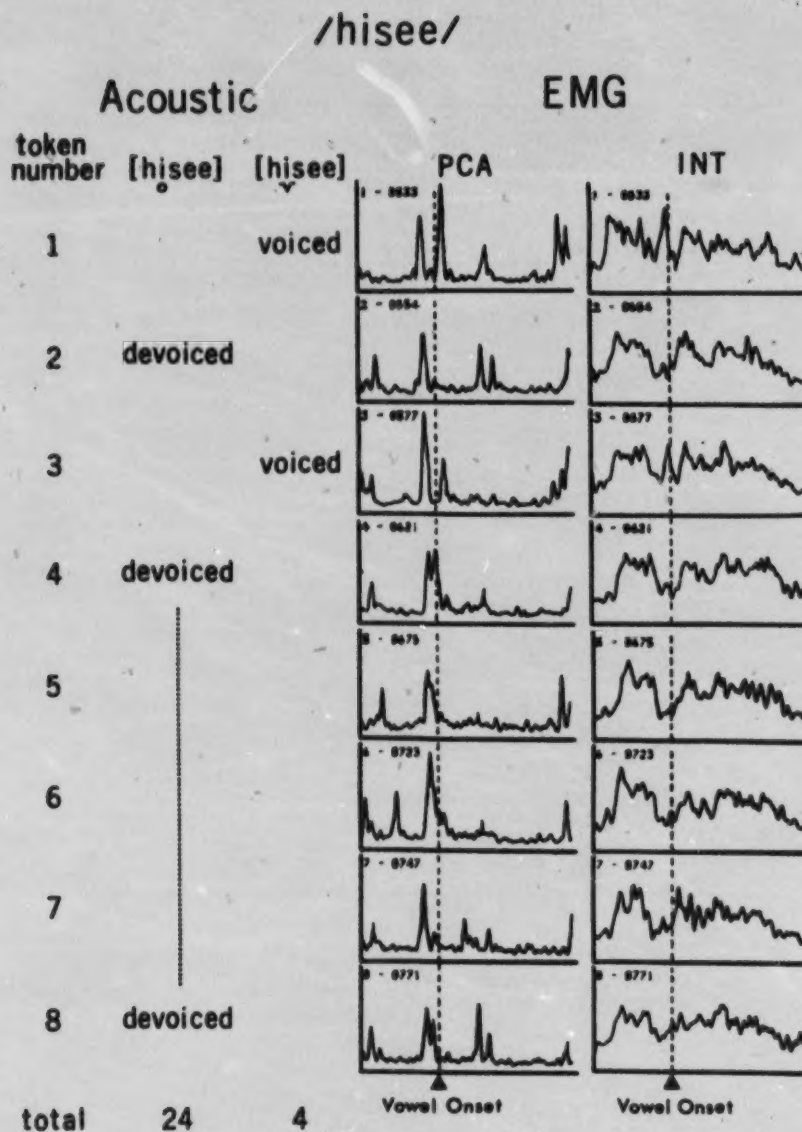


Figure 1. Sample of acoustic judgments and corresponding EMG activity patterns of PCA and INT for one of the test words containing the devoiceable vowel /i/. See text for details.

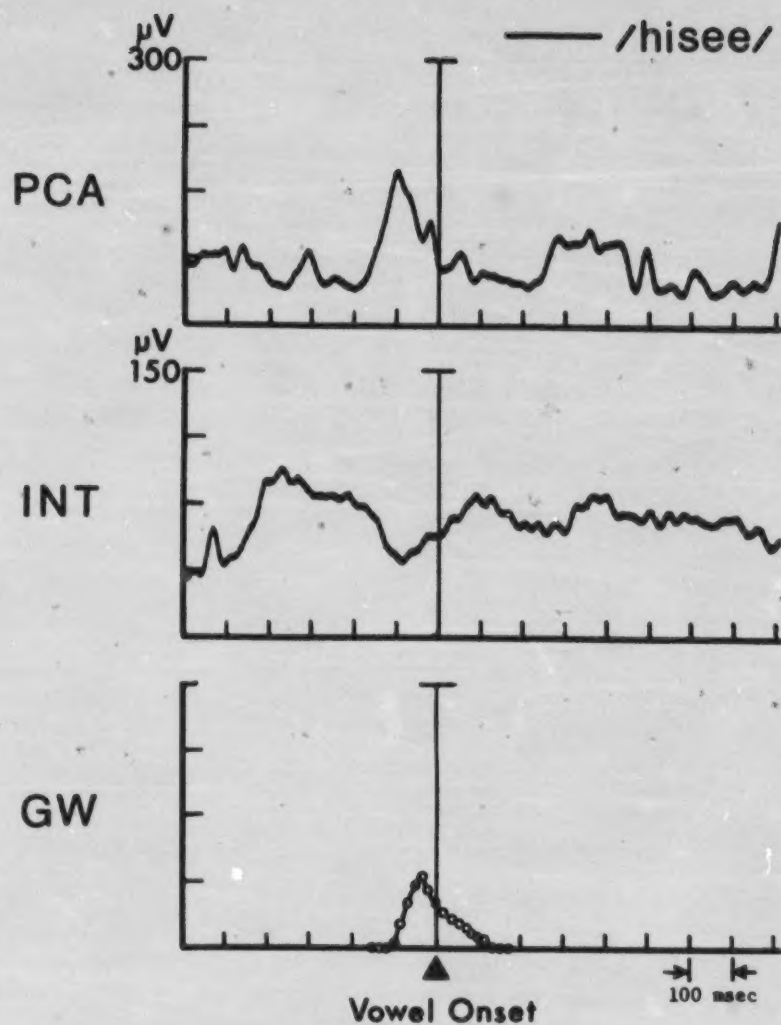


Figure 2. Averaged EMG curves of PCA and INT for all devoiced tokens of the same test word shown in Figure 1, together with a representative time course of glottal width.

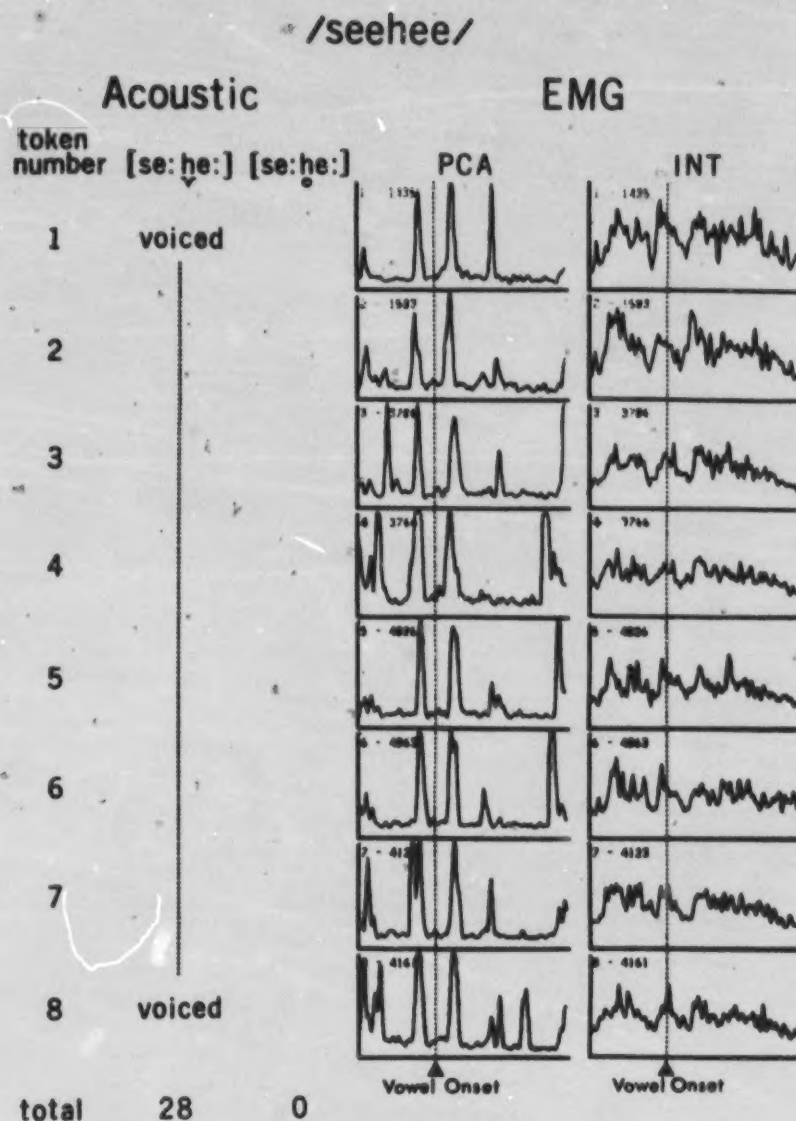


Figure 3. Sample of acoustic judgments and corresponding EMG activity patterns of PCA and INT for one of the test words containing intervocalic /h/. See text for details.

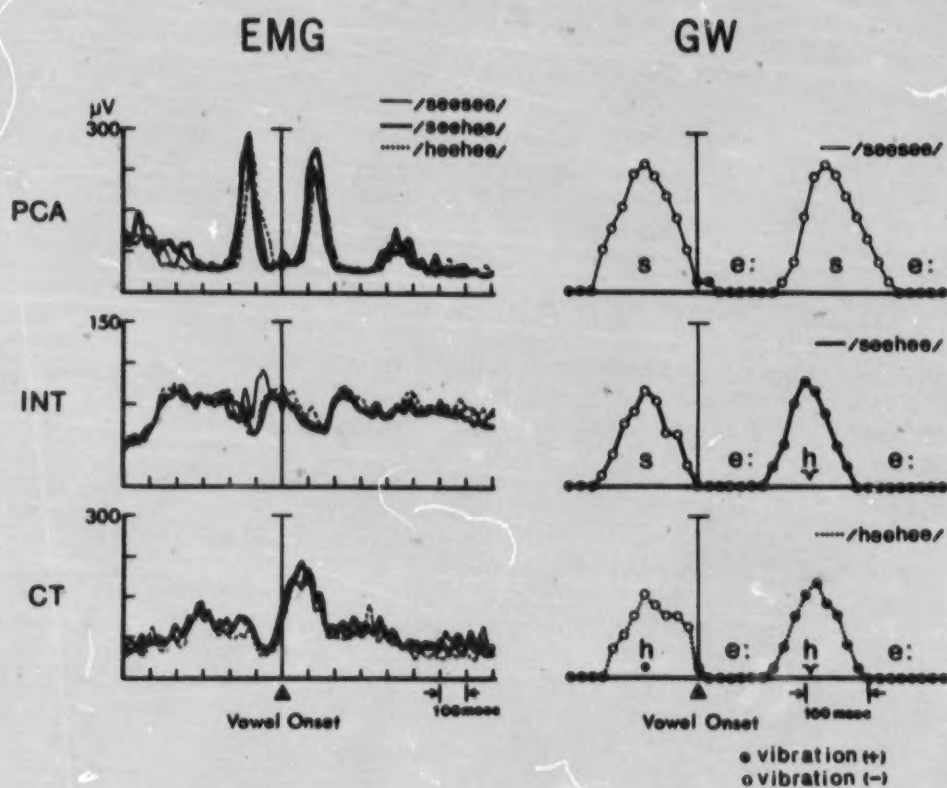


Figure 4. Superimposed averaged EMG curves of PCA and INT, and representative time courses of glottal width for three similar test words, including the one shown in Figure 3.

each token, the corresponding PCA and INT activity patterns, smoothed with a time constant of 35 msec, are illustrated at the right. The vertical dotted lines across the time axis indicate the acoustic onset of the vowel /i/, regardless of the voicing variation. For the two tokens judged as having voiced /i/, there are two clearly separate peaks near the line-up in their PCA curves. These peaks appear to correspond to the voiceless segments /h/ and /s/ respectively, while the INT curves show two dips intervened by a single peak, presumably corresponding to the voiced /i/. In contrast, for the tokens classified as devoiced, these activity patterns seem more variable, particularly in the PCA curves. Nevertheless, in general, PCA activity increases rather rapidly for the initial voiceless fricative /h/ and continues to increase for the following devoiced vowel segment with some time lead. The INT curve, on the other hand, shows one large continuous dip around the line-up point.

Figure 2 shows the averaged curves of the PCA and INT activity patterns, with a plot of glottal width as a function of time for a representative devoiced token included at the bottom. These curves, together with the EMG activity patterns of the single tokens in Figure 1, demonstrate that the activation of the PCA with the suppression of the INT, which is quite different from the pattern for the voiced tokens, is responsible for this vowel devoicing phenomenon, causing the glottis to remain open throughout the production of the devoiced vowel. It also means that, when the vowel /i/ does become devoiced, the neural command to these muscles differs for the same vowels. These findings were supported by comparable results drawn from the data for other devoiceable utterance types as well.

Figure 3 illustrates the acoustic judgments and EMG patterns for some tokens of another utterance type /seehee/, where the intervocalic /h/ is often accompanied by vocal fold vibration. In fact, as was shown in Table 1, all tokens were judged to be voiced in this sample. The EMG patterns of the PCA and INT seem more consistent than those for /hisee/, and show two almost identically shaped peaks in the PCA curves corresponding to the initial /s/ and the intervocalic voiced /h/ respectively, while the INT always shows nearly inverse activity, characterized by one peak for the interconsonantal vowel segment.

Figure 4 compares the averaged EMG activities for similar utterance types. The thick lines correspond to the test word /seehee/ mentioned previously, while the thin and dotted lines represent the utterances modified by changing the position of the two fricatives. It should be mentioned here that the word initial /h/ in /heehee/ was not accompanied by vocal fold vibration, although the intervocalic /h/ in /seehee/ or /heehee/ always was. Of course, the vibration did not occur for the voiceless fricative /s/ regardless of its position. In spite of these facts, the similarity in the PCA, INT and CT activity patterns among these three utterance types is obvious. The almost identical curves of the glottal width change as a function of time, shown at the right, are appropriate to confirm that the reciprocal activity patterns of PCA and INT can be accounted for in terms of the time course of the glottal opening and closing gesture rather than directly for the presence or absence of the vocal fold vibration in those particular cases.

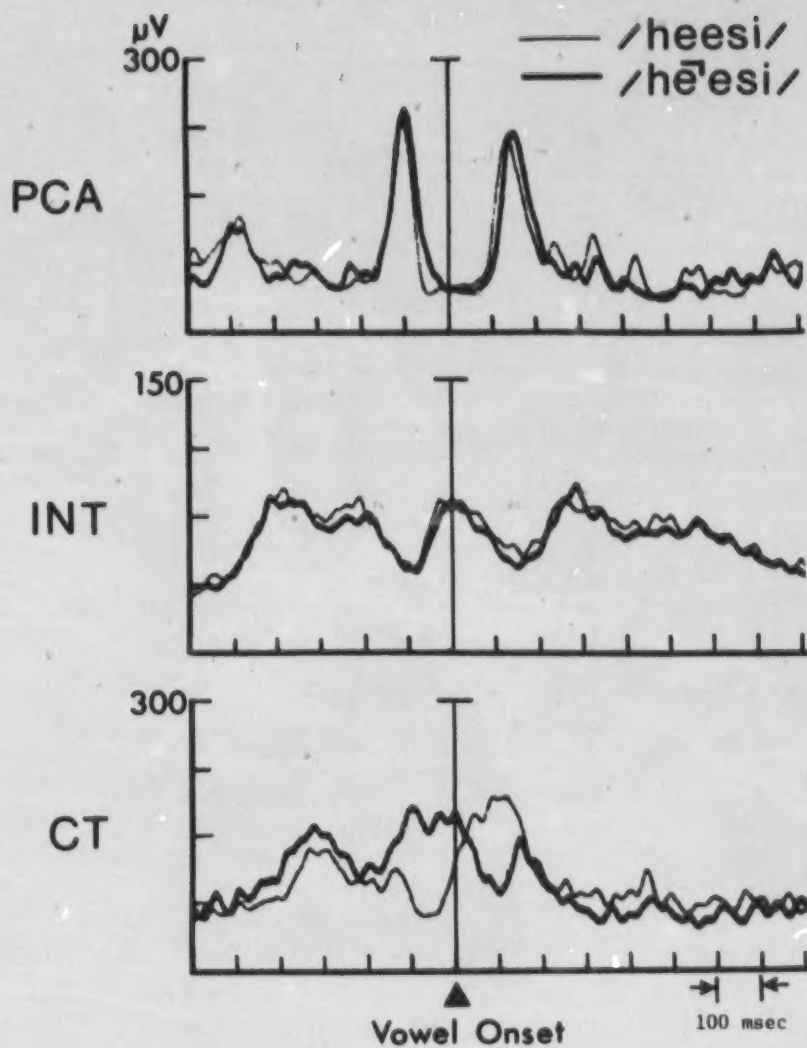
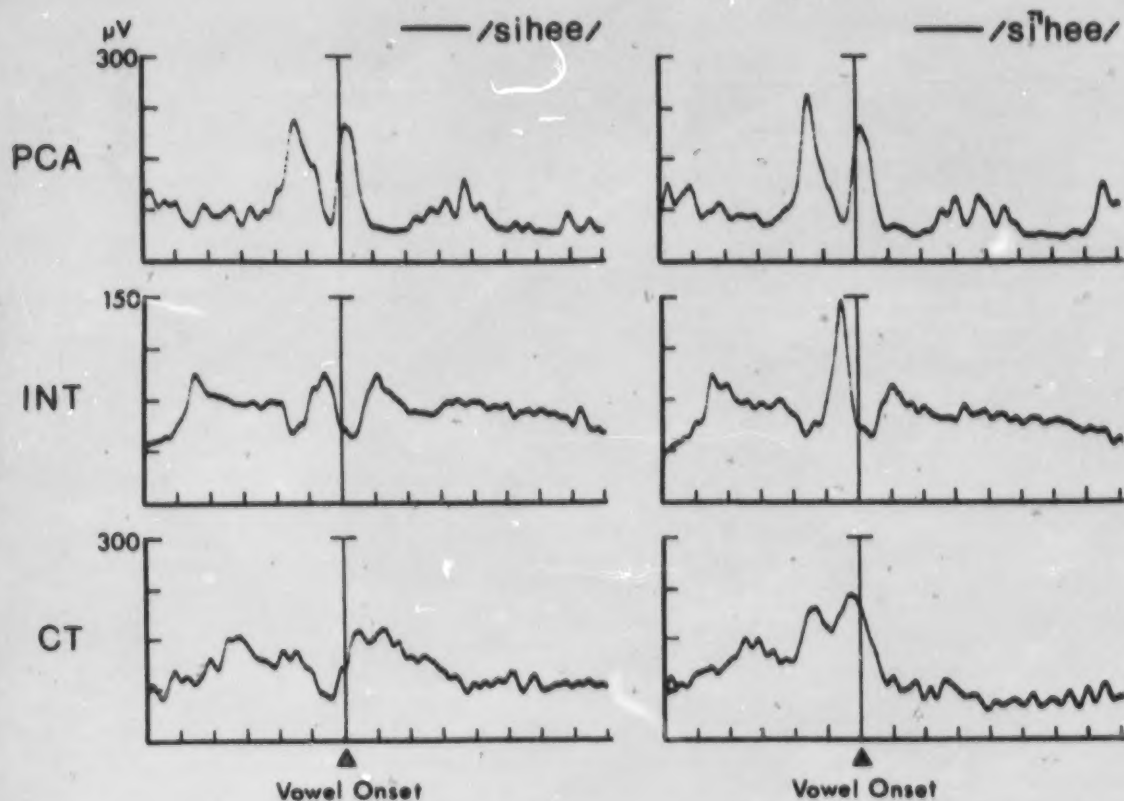


Figure 5. Averaged EMG curves of PCA, INT and CT for pair of the words /heesi/ and /he'esi/, which differ in accent type.

Figure 6. Averaged EMG curves of PCA, INT and CT for pair of words /sihee/ and /sɪ'hee/, either of which contains both devoiced vowel /i/ and intervocalic /h/. These curves were made with calculation for all tokens consequently uttered with fully voiced /i/ preceding voiced /h/, regardless of accent type.



As was shown in Table 1, it is commonly observed that the presence or absence of accent kernel in Japanese affects the devoicing of the high vowel and the following /h/, if present (e.g., Kindaichi, 1958; Han, 1962b; Sakurai, 1966). Before paying close attention to this interaction between the phonological pitch accent and the phonetic variation, the effect of the word accent on the phonetic manifestation of the voicing distinction is examined.

Figure 5 shows the EMG signals for two words consisting of the same phoneme sequence /heesi/, where the accent type varies. The PCA and INT patterns are almost identical and the time course of glottal opening was also found comparable under fiberoptic observation, which is not included here. Moreover, another pair of the words with /s/ instead of the initial /h/ was quite similar to the ones in this graph. In this regard, it is conceivable that the gross opening and closing gesture of the glottis and the temporal course of the PCA and INT activities for voiceless fricatives in Japanese are not significantly affected by the accent kernel, although the distinction of the accent kernel is clearly demonstrated by the activity patterns of the CT. Nevertheless, Figure 6 presents another piece of evidence related to the shift of devoicing caused by the accent distinction.

This graph shows the averaged EMG patterns for the phoneme sequence /sihee/ with accent type varying. It should be noted here that these curves were calculated for the subset groups where each token was eventually produced with fully voiced /i/ followed by voiced /h/, excluding the unvoiced cases. In other words, the segmental phonetic transcription for both groups is identical, allowing for the difference in pitch contour. One of the interesting findings is that the EMG patterns for the those muscles, particularly in the INT curves, appear to be significantly different, although the temporal course of the glottal width (not shown) is comparable. Specifically, the activity curve of INT for the segment /i/ without accent kernel, shown at the left, seems quite similar to that for the following vowel segments, while the right INT curve for the segment /i/ with accent kernel demonstrates a high and sharp peak compared to the following vowel. Furthermore, at least in this pair, that PCA activity for the voiceless consonant /s/ in the accented mora seems a little higher than for the other. These tendencies hold true for another pair of words having /s/ instead of /h/, i.e., /sisee/ vs. /sɪsee/ in spite of the absence of vocal fold vibration for the intervocalic fricative after the fully voiced /i/ or /i/. These results suggest that the accent command is manifested in strong activity of INT for the nuclear high vowel, presumably with slightly higher activity of the PCA for the preceding voiceless fricative, in the particular environment where this vowel is surrounded by voiceless consonants and consequently a candidate for vowel devoicing. The lesser probability of vowel devoicing for the accented nuclear vowel in such a context might be attributed to higher activity of INT. In other words, it is plausible that such a neurophysiological basis for definite closure of the glottis for the accented nuclear vowel is related to the effective manifestation of the presence of the accent kernel, which is usually realized by an abrupt drop in the pitch contour.

DISCUSSION

There are several experiments directed towards understanding the underlying physiological basis for allophonic variation in Japanese, mainly focused

on the vowel devoicing phenomenon by use of fiberoptic and EMG techniques (Sawashima, 1971; Hirose, 1971b; Weltzman, Sawashima, Hirose, & Ushijima, 1976). The present results appear to be generally in good agreement with those studies. Whenever the vowel devoicing is foregone, PCA and INT show EMG activity patterns clearly different from those for voiced tokens in a given phonological context. Thus, the devoicing phenomenon may be concluded to be an optional variation that has presumably originated at a higher level than the EMG signals in the speech production processes. Therefore, one could identify at least typical instances of vowel devoicing by inspection of EMG activity patterns only.

The present data also suggest, however, that there are considerable variations at the EMG level, particularly in the PCA curves, within the same devoiced group. In this connection, it is still difficult to conclude that this phonetic variation in voicing is a sort of binary adjustment predestined at or above the EMG level. In other words, there still remains the question whether these auditorily explicit two-way allophones are based on two different articulatory programming patterns at some neural level, or on the mere non-linear effect of rather wide fluctuations of the EMG potential. If the latter is the case, the significance of the averaged curve, especially for the devoiced group, should be reconsidered.

Moreover, the situation seems the more complicated, since the glottal opening gesture is reported to be quite distinct for these two groups. That is, the devoiced group is usually produced with a wide and single peaked opening gesture of the glottis throughout the vowel segment, while the other is definitely accompanied by a tight closure, as far as the published data are concerned. Although there are some articles implying inter-speaker variations of the time course of the glottal opening during the devoiced vowel segment (e.g., Sawashima, 1969; Sawashima, Hirose, & Yoshioka, 1978), there is no systematic study that suggests an analogous system at this level. It also means that the details of the configurational conversion from these responsible muscle contractions to the glottal shape are still unclear, particularly in quantitative terms. In this respect, it may be an alternative way to observe the activity patterns for a considerable number of single tokens of same utterance types, paying special attention to those critical cases.

As for the /h/ voicing phenomenon, which appears simply as the reverse phonetic situation to that of the vowel devoicing, the present data show that the EMG patterns and the changes in glottal aperture are essentially the same for all productions of /h/ regardless of the allophonic variation in voicing. It also means that the /h/ voicing does occur while the glottis remains as wide as for voiceless /h/. Furthermore, these patterns are found comparable to those for the phoneme sequence with /s/ replacing /h/, in spite of the fact that vocal fold vibration is never present for the fricative concerned in the latter cases. Thus, it may be that the glottal adjustments for the phonemic voiceless fricatives, such as /h/ and /s/, in Japanese are almost identical in terms of the gross opening and closing gesture of the glottis as well as the muscular control, although the participation of other laryngeal muscles including the thyroarytenoid and the lateral cricoarytenoid should be taken into account for a full description.

At the same time, it must be taken into account that the /h/ voicing is one of the exceptional cases, since other voiced sounds, such as vowels, voiced unaspirated stops, and other phonemic voiced fricatives are usually produced with more or less suppressed PCA activity supplemented by the increased INT activity, resulting in tight or loose closure, or at most slight opening of the glottis during these segments. It is also noted that the vocal fold vibration for this particular allophone in Japanese is observed as the tiny movement at the edges of the membranous portions, which is clearly differentiated from other phonologically relevant excitation under fiberoptic observation. Thus, the present results should be interpreted as entirely specific to this /h/ voicing phenomenon in this language.

In spite of these peculiarities in relation to other voiced sounds, it is an undeniable fact that /h/ in an intervocalic position is accompanied by at least a sort of vocal fold vibration. Furthermore, the quasi-periodicities are quite constantly detected in such environments. Thus, what enables the vocal folds to vibrate during the intervocalic /h/ production, in the face of the presumably disadvantageous situation of a widely separated glottis, should be found among some other physical condition(s) at the level of the glottis, including aerodynamic factors. Therefore, further research should be extended along this line to gain a more precise picture of the /h/ voicing phenomenon in Japanese.

REFERENCES

- Benguerel, A. P., Hirose, H., Sawashima, M., & Ushijima, T. Laryngeal control in French stop production: A fiberoptic, acoustic electromyographic study. *Folia Phoniatrica*, 1978, 30, 175-198.
- Dixit, R. P. Neuromuscular aspects of laryngeal control: with special reference to Hindi. Unpublished doctoral dissertation, University of Texas, 1975.
- Fischer-Jørgensen, E., & Hirose, H. A preliminary electromyographic study of labial and laryngeal muscles in Danish stop consonant production. Haskins Laboratories Status Report on Speech Research, 1974, SR-39/40, 231-254.
- Fujimura, O. Hatsuon no katei. *Japan Journal of Logopedics and Phoniatrics*, 1971, 12, 11-21.
- Han, M. S. Unvoicing of vowels in Japanese. *Study of Sound*, 1962, 10, 81-100. (a)
- Han, M. S. Japanese phonology: an analysis based on sound spectrograms. Tokyo: Kenkyusha, 1962. (b)
- Hattori, S. Onseigaku. Tokyo: Iwanamishoten, 1951.
- Hirano, M., & Ohala, J. Use of hooked-wire electrodes for electromyography of the intrinsic laryngeal muscles. *Journal of Speech and Hearing Research*, 1969, 12, 362-373.
- Hirose, H. Electromyography of the articulatory muscles: Current instrumentation and technique. Haskins Laboratories Status Report on Speech Research, 1971, SR-25/26, 73-85. (a)
- Hirose, H. The activity of the adductor laryngeal muscles in respect to vowel devoicing in Japanese. *Phonetica*, 1971, 23, 156-170. (b)
- Hirose, H. Posterior cricoarytenoid as a speech muscle. *Annals of Otology, Rhinology and Laryngology*, 1976, 88, 334-343.
- Hirose, H., & Gay, T. The activity of the intrinsic laryngeal muscles in

- voicing control: An electromyographic study. Phonetica, 1972, 25, 140-164.
- Hirose, H., Gay, T., Strome, M., & Sawashima, M. Electrode insertion technique for laryngeal electromyography. Journal of the Acoustical Society of America, 1971, 50, 1449-1450.
- Iwata, R., & Hirose, H. Fiberoptic acoustic studies of Mandarin stops and affricates. Research Institute of Logopedics and Phoniatrics, University of Tokyo, Annual Bulletin, 1976, 10, 47-60.
- Kagaya, R. A fiberoptic and acoustic study of the Korean stops, affricates and fricatives. Journal of Phonetics, 1974, 2, 161-180.
- Kagaya, R., & Hirose, H. Fiberoptic, electromyographic and acoustic analysis of Hindi stop consonants. Research Institute of Logopedics and Phoniatrics, University of Tokyo, Annual Bulletin, 1975, 9, 27-46.
- Kewley-Port, D. Computer processing of EMG signals at Haskins Laboratories. Haskins Laboratories Status Report on Speech Research, 1973, SR-33, 173-183.
- Kewley-Port, D. An experimental evaluation of the EMG data processing system: Time constant choice for digital integrations. Haskins Laboratories Status Report on Speech Research, 1974, SR-37/38, 65-72.
- Kewley-Port, D. EMG signal processing for speech research. Haskins Laboratories Status Report on Speech Research, 1977, SR-50, 123-146.
- Kindaichi, H. Meikai nihongo akusento jiten. Tokyo: Sanseido, 1958.
- Lisker, L., Abramson, A., Cooper, F. S., & Schvay, M. H. Transilluminations of the larynx in running speech. Journal of the Acoustical Society of America, 1969, 45, 1544-1546.
- Sakurai, S. Boin no museika. In Nihon Hoso Kyokai (Ed.) Nihongo hatsuon akusento jiten. Tokyo: NHK, 1966.
- Sawashima, M. Devoiced syllables in Japanese: a preliminary study by photoelectric glottography. Research Institute of Logopedics and Phoniatrics, University of Tokyo, Annual Bulletin, 1969, 3, 35-41.
- Sawashima, M. Glottal adjustments for English obstruents. Haskins Laboratories Status Report on Speech Research, 1970, SR-21/22, 187-200.
- Sawashima, M. Devoicing of vowels. Research Institute of Logopedics and Phoniatrics, University of Tokyo, Annual Bulletin, 1971, 5, 7-13.
- Sawashima, M. Hatsuonji no koto chosetsu. In S. Hiki (Ed.), Onseikagaku. Tokyo: Tokyo University Press, 1973.
- Sawashima, M. Current instrumentation and technique for observing speech organs. Technocrat, 1977, 9-4, 19-26.
- Sawashima, M., & Hirose, H. A new laryngoscopic technique by use of fiberoptics. Journal of Acoustical Society of America, 1968, 43, 168-169.
- Sawashima, M., Hirose, H., & Yoshioka, H. Abductor (PCA) and adductor (INT) muscles of the larynx in voiceless sound production. Research Institute of Logopedics and Phoniatrics, University of Tokyo, Annual Bulletin, 1978, 12, 53-60.
- Sawashima, M., & Miyazaki, S. Glottal opening for Japanese voiceless consonants. Research Institute of Logopedics and Phoniatrics, University of Tokyo, Annual Bulletin, 1973, 7, 1-9.
- Weitzman, R. S., Sawashima, M., Hirose, H., & Ushijima, T. Devoiced and whispered vowels in Japanese. Research Institute of Logopedics and Phoniatrics, University of Tokyo, Annual Bulletin, 1976, 10, 61-79.

ON THE COORDINATION OF TWO-HANDED MOVEMENTS*

J. A. Scott Kelso,⁺ Dan L. Southard,⁺⁺ and David Goodman

Abstract. In a set of three experiments, we show that after an auditory "go" signal, subjects simultaneously initiate and terminate two-handed movements to targets of widely disparate difficulty. This is the case when the movements required are (a) lateral and away from the midline of the body (Experiment 1), (b) towards the midline of the body (Experiment 2), and (c) in the forward direction away from the body midline (Experiment 3). Kinematic data obtained from high-speed cinematography (200 frames/sec) point to a tight coordinative coupling between the hands. Although the hands move at entirely different speeds to different points in space, times to peak velocity and acceleration are almost perfectly synchronous. We promote the viewpoint that the brain produces simultaneity of action as the optimal solution for the two-handed task by organizing functional groupings of muscles--coordinative structures--that are constrained to act as a single unit.

INTRODUCTION

Recent theoretical development in motor behavior has focused to a considerable degree on the issue of whether movements are under closed-loop (feedback) or open-loop (programmed) control (Adams, 1971, 1977; Schmidt, 1975). Much of the data has been generated from linear positioning tasks involving the use of a single limb. In contrast, little is known about the principles governing interlimb coordination, even though much of human movement involves the coordinated participation of both hands and hence the concerted operation of the cerebral hemispheres (Luria, 1973). Part of the reason for this state of affairs may be that coordination does not lend itself easily to quantification. Rather, we seem content to rely on anecdotal evidence for insight into such problems.¹

*Appeared in Journal of Experimental Psychology: Human Perception and Performance, 1979, 5(2), 229-238. Experiment 1 of the present paper is also reported in Science, 1979, 203, 1029-1031.

⁺Also University of Connecticut, Storrs.

⁺⁺Presently at the University of Iowa.

Acknowledgment: This research was supported by Biomedical Research Support Grant FR-07035 from the General Support Branch, the National Institutes of Health, and by U.S. Public Health Service Grant 1 R01 AM20231-01 (National Institute of Arthritis, Metabolism, and Digestive Diseases). We thank Mike Linton and Pat Briley for technical assistance, Carol Putnam for cinematographical analysis, and Phil Striegel for running subjects in Experiment 1. The work was conducted at the University of Iowa, Iowa City.

In this paper we present three experiments on a behavioral task that involves coordination of the upper limbs. Combined with high-speed cinematographic movement analysis, the findings elucidate the mode of control utilized by subjects when faced with a task that places different movement demands on each hand. Our question was a simple one. Suppose an individual is asked to produce movements of the upper limbs to targets each of which varies in amplitude and precision requirements; How will he or she respond? A relationship between movement duration, movement amplitude, and target-demands, formulated some time ago by Fitts (1954), allows us to examine this question experimentally. The equation relating the foregoing parameters is known as Fitt's Law, in which movement time = $a + b \log_2 (2A/W)$, where a and b are constants, A is the amplitude of the movement, and W is the width of the target. The key aspect of this formulation is that movement time depends on the ratio of movement amplitude to movement precision. Thus, the movement time for a 4-cm movement to a .5-cm target width (8:1 ratio) is practically identical to an 8-cm movement to a 1-cm target.

Consider a one-handed movement condition in which the target size is large and the amplitude short (termed easy), relative to a condition in which the target size is small and the movement amplitude is long (termed difficult). Movement time in the former case will obviously be shorter in duration. But what happens when these conditions are combined for both hands? Does the hand producing a short movement to an easy target arrive much earlier than the more difficult condition or are the movements initiated and terminated simultaneously?

A pilot experiment was conducted to examine this question. Ten subjects performed single- and two-handed movements (involving extension of the wrist-forearm linkage) of equal and varying difficulty as quickly and as accurately as possible, following an auditory stimulus. A major finding was that movement times for the easy task under combined conditions (that is, easy for one limb and difficult for the other) were nearly doubled compared with single limb counterparts and conditions where both hands performed the easy task. The hand moving to the easy target under combined conditions therefore appeared to wait for the hand traveling to the difficult target so that they could strike together.² This finding indicates that in spite of differences in target demands and movement length between each hand, response duration appears to be held constant. Duration, then, could be interpreted as a major parameter in the program for two-handed movements. One of the drawbacks of the pilot experiment was that subjects were instructed, at the onset of an auditory signal, to leave the home keys simultaneously. Indeed, trials in which reaction time differences between the hands were greater than 15 msec were excluded. Although this criterion was exceeded on a very small proportion of trials, we felt that the emphasis on simultaneous reaction time may have biased subjects to also terminate the movements simultaneously. The procedure in Experiment 1, therefore, was simply to instruct the subjects to strike the designated targets as quickly and as accurately as possible, without any reference to reaction time simultaneity. We felt that removal of this potential bias would provide a clearer picture of how the limbs perform under combined conditions.

EXPERIMENT 1

Method

Subjects. The subjects were 12 right-handed volunteers ranging in age between 18 and 25 years. One subject's results were excluded from the data analysis because limited peripheral vision prevented his performing the task in certain movement conditions without an exceptional number of errors. Although the subjects were not paid individually for their participation, a five dollar bonus was awarded to the most accurate subject with the best overall response times (i.e., combined reaction times and movement times).

Apparatus. The apparatus consisted of a Plexiglas base (76 cm in length, 16 cm wide, and .8 cm thick) mounted on a standard table (75 cm high) such that the long edge of the base was parallel to the front edge of the table. Two normally closed momentary-contact switches (Cherry keyboard switch, Model #M62-0900), centered 4.5 cm apart, served as the home keys. The base was constructed so that two hinged masonite targets could be positioned along the longitudinal center line of the base, anywhere from 2 cm to 32 cm in distance from the home keys. Two target widths were used: The "easy" target was 7.2 cm wide and the "difficult" target was 3.6 cm wide. These were located at either a short distance (6 cm) or a long distance (24 cm) from the home keys. A single target was used in single-hand conditions and two targets in the double-hand condition, allowing all combinations of target width and target distance to be utilized. A red light-emitting diode served as the warning light and the sound from a Minisonalert provided the stimulus to move. These were mounted on a 50 cm x 15 cm board centered 10 cm behind the apparatus, directly in front of the subject. The onsets of warning light and stimulus tone were controlled by a PDP8/A computer that also collected reaction times, movement times, and total response times.

Task. The subject's task was to move his or her index fingers from the home keys to the targets as fast and as accurately as possible after receiving the auditory stimulus from the Minisonalert. For single-hand conditions, the subject depressed the left home key with the left index finger or the right home key with the right index finger, and, on receiving the stimulus to move, proceeded to the designated target, touching it only with the index finger. For two-handed conditions, the subject depressed both home keys with the index fingers and proceeded to hit the respective targets following the onset of the auditory stimulus. All movements from the home keys to the targets were lateral.

Procedure. Eight experimental conditions were used, which varied depending on (a) whether a single- or two-handed movement was required, (b) whether the target was easy or difficult and, (c) whether the movement was of short or long amplitude. The nature of the task was explained to the subjects and the instructions emphasized both speed and accuracy in striking the target(s). When the experimenter was certain that the subject understood the instructions, all eight conditions were performed by the subject. Each condition consisted of 25 trials with a 5-sec intertrial interval and a 1- to 3-sec variable foreperiod between the warning light and the stimulus to move. Only the last 20 trials of each condition were used in the data analysis; the first 5 trials served as familiarization. When each trial block was completed, the

Total Response Time	Movement Time	Reaction Time	Left Target	Home Keys	Right Target	Reaction Time	Movement Time	Total Response Time
				• •	1 □	218	159	377
371	151	220	2 □	• •				
287	82	205	4 □	• •				
				• •	3 □	218	78	296
308	89	219	6 □	• •	5 □	224	85	309
403	186	237	8 □	• •	7 □	240	189	409
383	165	238	10 □	• •	9 □	246	133	379
383	140	243	12 □	• •	11 □	240	158	398

Figure 1. Mean reaction time, movement time, and total response times (in msec) for single- and two-handed movements directed away from the midline of the body. For actual dimensions of the targets and their distance from the home keys, refer to the text.

177

subject was given a 3-min break during which the experimenter re-arranged the targets in preparation for the next movement condition. All movements to targets were monitored by the experimenter. If the subject missed the target or hit the target with anything other than the index finger, that trial was excluded from the data analysis. Furthermore, reaction times greater than 600 msec or less than 90 msec and movement times greater than 600 msec and less than 30 msec were also excluded.

Design

A within-subjects design was used with all 11 subjects performing in all experimental conditions, whose order was randomized. From the 20 trials in each condition, mean reaction time, movement time, and total response time were computed for each hand. There were four single-handed and four two-handed conditions, making a total of 12 separate means for each subject and for each dependent variable. Preplanned contrasts using Dunn's procedure (Kirk, 1968, p. 79) were carried out on the means of interest.

Results and Discussion

The mean reaction times, movement times, and total response times are shown for each condition in Figure 1. Given the current debate regarding the use of simple versus choice reaction time as a reflection of the time it takes to select and prepare or "program" upcoming motor responses (e.g., Klapp, 1977; Sternberg, Monsell, Knoll, & Wright, 1978), we prefer not to interpret our results within that theoretical framework. Our chief concern was whether subjects initiated and terminated movements simultaneously, especially under conditions where the task demands were different for each hand.

No significant hand differences in reaction time were found ($p > .05$). More interestingly, subjects appeared to initiate hand movements in paired conditions virtually simultaneously. This is apparent in Figure 1 where the largest difference between left- and right-hand reaction times is 8 msec (9 and 10). Thus, subjects left the home keys together even in the absence of instructions to do so. The average within-subject correlation between left and right hands in paired conditions was also extremely high (range .95 to .97), further supporting the simultaneity of initiation.

As can be seen in Figure 1, single-handed movement times for the easy task (3 and 4) are much faster than their difficult counterparts (1 and 2) as Fitts' Law predicts ($p < .05$). This effect is also evident when examining two-handed movements (5 and 6 versus 7 and 8, $p < .05$). Movement times for single- and two-handed movements of the same difficulty are not significantly different ($p > .05$). However, when the task demands are varied for each hand, movement times for the easy task (9 and 12) are significantly elevated over paired easy conditions (5 and 6), $p < .01$. Clearly, the difficult task determines movement time in two-handed conditions.

The movement time data in Figure 1 also indicate that two-handed movements of equal difficulty are executed simultaneously (5 versus 6 and 7 versus 8). Furthermore, paired movements of varying difficulty are also executed virtually simultaneously. Movement times to the easy target (9 and 12) are only slightly faster than movement times to the difficult target (10

and 11). In fact, when total response times are considered, this difference (19 msec) is non-significant ($p > .05$).

The overall error rate across the eight experimental conditions was 8%. These ranged, as expected, from a small error rate in single-hand conditions (6%) to higher errors in two-handed difficult conditions (13%). The majority of these errors was due to the subject's missing the target or failing to strike the target with the designated finger.

The results of Experiment 1 essentially replicated those of the pilot experiment. The reaction time data strongly suggest that subjects initiated two-handed movements at the same time. Furthermore, paired movements to targets of equal or unequal difficulty were terminated simultaneously, as is evident in their corresponding movement times and total response times. Even though the task demands were quite different under combined conditions, the hands appear to perform in a unitary manner. One drawback to this conclusion is that the outcome of Experiment 1 may have arisen as a result of the targets being placed in the subject's peripheral vision. Thus, subjects may simply have attended to or monitored movement to the difficult target, leaving the contralateral hand to perform a subsidiary role.³ In Experiment 2 we wanted to check whether this was a necessary and sufficient condition for the apparent time dependence between the hands. The way we chose to confront this issue was to have both movements terminate in focal vision. To accomplish this, we simply interchanged targets with home keys so that the former were placed directly in front of the subject.

EXPERIMENT 2

Method

Subjects. The subjects were 12 student volunteers who had not participated in Experiment 1 or the pilot study. One subject's data were lost due to equipment malfunction.

Apparatus. The apparatus was similar in design to that used in Experiment 1, the only difference being that the position of the home keys and targets was interchanged. Thus, the targets were now directly in front of the subject and the home keys could be adjusted to different distances from the targets. The task therefore involved flexion primarily of the elbow joint towards the midline of the body. Target dimensions and movement amplitudes were the same as those in the previous experiment.

Procedure and Design. The procedures for Experiment 2 were identical to those of Experiment 1, except that subjects received only 20 trials per condition. The first five trials served as familiarization and were not included in the analysis. Preplanned comparisons were carried out on relevant mean reaction times, movement times, and total response times.

Results and Discussion

The mean reaction times, movement times, and total response times are shown for each condition in Figure 2. As in Experiment 1, no significant

Total Response Time	Movement Time	Reaction Time	Home Keys	Targets	Home Keys	Reaction Time	Movement Time	Total Response Time
				□1←—⊕		231	218	449
445	221	224	⊕→2□					
				□3←—⊕		228	140	368
368	140	229	⊕→4□					
385	150	235	⊕→6□	□5←—⊕		243	145	388
448	216	232	⊕→8□	□7←—⊕		237	220	457
451	213	238	⊕→10□	□9←—⊕		253	192	445
427	183	244	⊕→12□	□11←—⊕		238	209	447

Figure 2. Mean reaction time, movement time, and total response times (in msec) for single- and two-handed movements directed towards the midline of the body.

differences in reaction time were found ($p > .05$). The largest difference between the hands was 15 msec (9 and 10), which was not significant. That subjects' hands were leaving the home keys together is further supported by the high within-subject correlation between left and right hands (range .74 to .98).

The data again indicated the expected relationship between single-handed movements for the easy task (3 and 4) and the single-handed movements for the difficult task (1 and 2), with the easy task clearly resulting in faster movement times ($p < .01$). This effect was also evident in two-handed movements (5 and 6 versus 7 and 8, $p < .01$). Once again, the two-handed movements of equal difficulty (5 versus 6 and 7 versus 8) were executed simultaneously. As in Experiment 1, the difficult task appears to determine the movement time in two-handed conditions. The slight movement time advantage of the easy task (9 and 12) over the difficult task (10 and 11) in combined conditions disappears when total response time is considered ($p > .05$). The overall error rate across the eight experimental conditions was 1.8%.

The results of Experiments 1 and 2 revealed identical effects, in that simultaneity of initiation and termination occurred in all combined movement conditions. It should be noted that in both experiments the task involved symmetrical muscle groups resulting in movements in opposite directions. To further examine the generality of the simultaneity effect we employed a task that also involved symmetrical muscle groups, but that required movements in the same direction. Consider the case where the subject must produce two-handed movements of varying difficulty in the forward direction. An opportunity is afforded the subject to terminate the easy task before the difficult one. Thus, if both hands are initiated together and proceed forward at the same rate, the subject could feasibly strike the near target first, and the simultaneity effect would break down.

EXPERIMENT 3

Method

Subjects. The subjects were 12 student volunteers who did not participate in either of the previous studies.

Apparatus. The basic model of the apparatus remained consistent with Experiments 1 and 2. However, the equipment was altered so that movements could be made forward in the sagittal plane, rather than laterally. This was accomplished by having two identical pieces of Plexiglas (106 cm long, 7 cm wide, and .8 cm thick), each with a single home key and moveable and interchangeable targets. Target widths and distances from the home keys were the same as those used in Experiments 1 and 2. The two pieces of apparatus were positioned parallel to each other, extending forward from the seated subject. The warning display and auditory stimulus setup was identical to the two previous experiments. The onsets of warning light and stimulus tone were controlled by a PDP 8/A computer, which also collected reaction times, movement times, and total response times.

Procedure and Design

The procedures and design for Experiment 3 were identical to those of Experiment 2, with all subjects performing in all eight experimental conditions in a randomized order.

Results and Discussion

The mean reaction times, movement times, and total response times are shown for each condition in Figure 3. The subjects initiated movements together, as indicated by the null effect of right versus left hand in the paired conditions and the high within-subject correlation (range .82 to .98).

The easy versus difficult task manipulation was effective as indicated by the longer movement times to the far target, both in the single-hand condition (3 and 4 versus 1 and 2, $p < .01$) and in paired conditions of the same difficulty (5 and 6 versus 7 and 8, $p < .01$). The difficult task once again exerted a major influence on the movement time in combined conditions as evident in the increase in movement time of the easy hand when the contralateral hand performs the more difficult task. The major finding of simultaneity once again appeared with the slight movement time advantage of the easy task (10 and 11) over the difficult task (9 and 12) being further reduced when one considers total response time (mean difference 14 msec, $p > .05$). The overall error rate across the eight experimental conditions was 1.0%.

GENERAL DISCUSSION

There is a remarkable consistency in the pattern of results across the three experiments. First, notice that movement times for the so-called difficult task in single-hand conditions are greater than for the easy task. Second, the easy-difficult difference carries over to two-handed movements when the task is the same for each hand. But most interesting is the finding that movement times for paired movements of unequal difficulty are virtually identical. When total response times are considered, any difference in termination between the hands is greatly reduced. This set of findings cannot be attributed to a peripheral vision problem (see Experiment 2), nor to the fact that in Experiments 1 and 2 the hands always move in opposite directions. When subjects are afforded the opportunity to break down the apparent time dependence between the hands in Experiment 3, they do not take it. In all three experiments, then, subjects initiate and terminate symmetrical movements of the hands to different points in space virtually simultaneously. A key issue for the present paper concerns whether the limbs are controlled as separate units in the easy-difficult case or, conversely, whether they are constrained to act as a single unit. More specifically, do the central commands prescribe the details of the intended movements for each hand or, alternatively, are central commands referred to functional groupings of muscles that operate fairly autonomously to produce simultaneity of action?

It seems quite tempting, for example, to interpret the present data in terms of a central program specifying different commands for each limb. The parameter remaining constant in this case--movement duration--might be viewed as "setting the limits" for the commands generated. Indeed, this is not an unreasonable position, for there is ample evidence from reaction time/movement

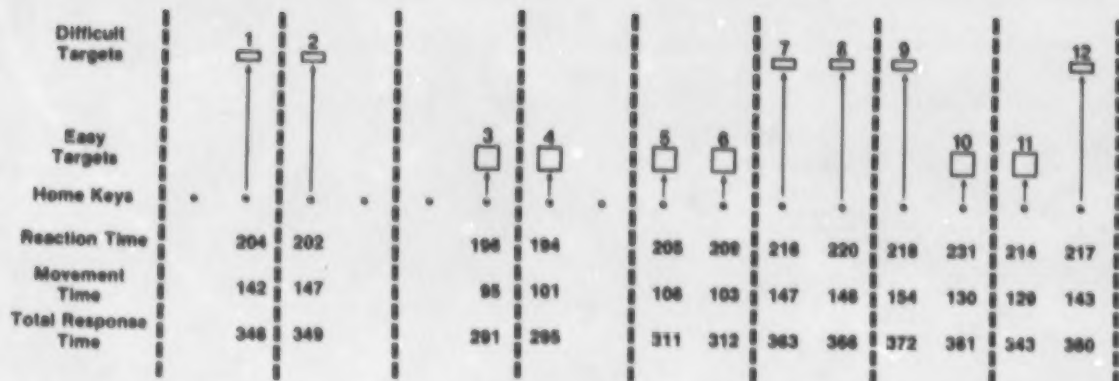


Figure 3. Mean reaction time, movement time, and total response times (in msec) for single- and two-handed movements in the forward direction.

time studies that duration is a major variable influencing the programming process (Kerr, 1978, for a review). Furthermore, recent neurophysiological data suggest that the duration parameter is centrally preprogrammed (Brooks, 1974; Koslovskaya, Atkin, Horvath, Thomas, & Brooks, 1974). When the location of mechanical stops was altered unbeknownst to monkeys producing rapid alternating elbow movements, they nevertheless maintained movement duration constant. Thus, rather than oscillating between the stops as quickly as possible, they exerted force against the newly placed stops, keeping the originally-learned rhythmic pattern stable.

But a rather different mode of control may be suggested from Bernstein's (1967) original work and subsequent research on activities such as locomotion (see Boylls, 1975; Grillner, 1975; Shik & Orlovsky, 1976, for reviews) and respiration (Gurfinkel', Kots, Pal'tsev, & Fel'dman, 1971). Movements are viewed as centrally programmed, not in terms of individual muscle contractions but rather according to muscle linkages. A linkage is defined as a group of muscles whose activities covary as a result of shared efferent or afferent signals (Boylls, 1975). For example, extensive studies on locomotion in animals reveal that movements are organized in terms of basic flexor and extensor linkages--spinal locomotor automatisms (Shik & Orlovsky, 1976)--involving both proximal and distal joints.

This basic mode of motor organization is revealed in an experiment--somewhat analogous to the present studies--performed by Kulagin and Shik (1970) on mesencephalic cats running on a treadmill at two different speeds. In this situation the movements of the two sides of the body are different just as they are in normal activities such as turning or circling. Although the speeds of symmetrical limbs were obviously different and took the form of a strict alternation pattern, the duration of the step cycle remained constant. This was achieved by lengthening the stance phase and shortening the swing phase on the slower belt, with a concomitant shortening of the stance phase and lengthening of the swing phase on the faster belt.⁴ It appears that a low level mechanism is involved in this interaction between the two sides of the body, for an identical result occurs in the spinal animal (Grillner, 1975).

The picture of interlimb coordination that emerges from studies of this type is that the task of central signals is not to prescribe the details of the intended movement but rather to organize functional groupings of muscles--coordinative structures (Easton, 1972; Turvey, 1977)--in a relatively autonomous fashion. Viewed in light of the present experiments, this style of control argues that the brain sets the level of activity in low level automatisms based on the spatial demands of the task, but leaves them to generate the pattern of interlimb coordination seen in simultaneous movements. Indeed, we have data that suggest that in a task where the spatial demands vary on each side, the limbs are constrained to function as a single unit. High-speed cinematographic analysis (200 frames/sec) reveals that the limb moving to the easy target does not hover over the target or "wait" for its difficult counterpart, but moves at an entirely different speed. More importantly, as Figure 4 reveals, the limbs under easy-difficult target conditions reach peak velocity and peak acceleration at practically the same time during the movements. Thus, although the limbs move at different speeds, their velocity and acceleration patterns are nearly perfectly synchronous.

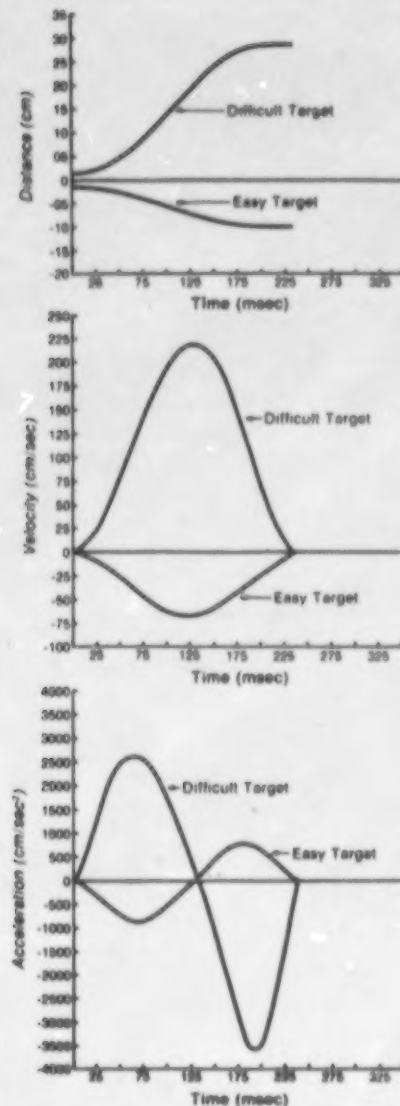


Figure 4. The pattern of displacement, velocity, and acceleration over time for two-handed movements of unequal difficulty obtained from single frame kinematic analysis (frame rate = 200 frames/sec). Over a series of six trials the mean time difference in peak velocities was 9 msec, while the mean difference between peak accelerations was 14 msec for positive acceleration and 4 msec for negative acceleration.

This suggests a strong interaction between the limbs and is not conducive to an independent programming view. The apparently fixed and reproducible interaction between the limbs seen in the present experiments to produce simultaneity of action may be viewed as the discovery of a coordinative structure or muscle linkage, a goal that has motivated much of the Russian work on motor control (e.g., Gurfinkel' et al., 1971). The notion that motor coordination involves a reduction of the degrees of freedom of the motor apparatus, first advanced by Bernstein (1967) and lately extended by Turvey (1977), requires the existence of low level coordinative structures that govern the interaction between limbs. Such collectives are not necessarily prefabricated, as Easton (1972) has argued in the case of reflexes. Rather, they are functional and may be marshalled temporarily and expressly for the purpose of accomplishing a particular behavioral goal.

This perspective on coordination raises numerous theoretical issues. Boylls (1975), for example, has discussed how the deployment of coordinative structures is parameterized. At one level is the structural prescription defined as a set of qualitative ratios of activities in the linked muscles, independent of absolute activity levels. On the other hand, the metrical prescription of a coordinative structure specifies the absolute level of activity in linked muscles. The latter may be viewed as a scalar quantity that multiplies the activities of all muscles in the linkage. Boylls argues, with respect to the anterior lobe of the cerebellum, that structural prescriptions are tuned by adjusting the relative amounts of activity distributed among descending tracts from the cerebellum, while metrical prescriptions are governed by the absolute activity levels in those tracts. This view receives strong support from Orlovskii's (1972) data showing that cerebellar stimulation during cat locomotion affects only the magnitude of muscle contraction, leaving unchanged both the period duration and the timings of periods relative to the cat cycle. This may be the principal characteristic of a coordinative structure. Namely, when a group of muscles is constrained to act as a unit, some temporal relationship is preserved invariantly over changes in the magnitude of activity (Turvey, Shaw, & Mace, 1978).

Our data on two-handed movements fit this theoretical perspective rather well. When the movement kinematics are examined, it is quite obvious that the magnitude of forces produced for each hand is different (see Figure 4). Thus, the equilibrium points for each hand may be preset and the neural output specified accordingly in terms of the magnitude of forces required (Bizzi & Polit, in press; Kelso, 1977). However, the underlying temporal structure remains invariant between the hands such that they preserve a synchronous relationship to each other. Hence, the metrical prescription (specified by the spatial parameters) is modulated for each hand, yet the structural prescription (the relative timing between the hands) remains invariant.

In conclusion, the present experiments represent an initial attack on a problem that has been largely ignored by motor behavior researchers, namely interlimb coordination. Consequently, apart from some recent theorizing of a preliminary nature (Turvey, 1977; Fowler & Turvey, 1978) formal theoretical development has been sadly lacking. We feel that the present behavioral paradigm, especially when combined with movement analysis techniques, has broad potential for examining coordination issues. Our data suggest that when the motor system is faced with controlling multiple degrees of freedom, as in

the two-handed task, it solves the problem optimally by constraining the limb musculature to act as a single unit. If this is so, then variables designed to influence one limb's moving to a spatial target (such as slowing the limb down or requiring a change in the limb's angle of projection) should have concomitant modulatory effects on the other limb. Of course, we do not claim that the performer cannot break down these restraints with practice. Many motor tasks require the hands to perform in an independent rather than tightly coupled manner. In the broader perspective, therefore, highly skilled performance might be viewed as a release from the type of temporal invariance exhibited in these experiments.

REFERENCES

- Adams, J. A. A closed-loop behavior theory of motor learning. Journal of Motor Behavior, 1971, 3(2), 111-149.
- Adams, J. A. Feedback theory of how joint receptors regulate the timing and positioning of a limb. Psychological Review, 1977, 84, 504-523.
- Bernstein, N. The co-ordination and regulation of movements. Oxford: Pergamon Press, 1967.
- Bizzi, E., & Polit, A. Processes controlling visually evoked movements. Neuropsychologia, in press.
- Boylls, C. C. A theory of cerebellar function with applications to locomotion. II. The relation of anterior lobe climbing fiber function to locomotor behavior in the cat. COINS Technical Report, 1975, 76(1). Department of Computer and Information Science, University of Massachusetts.
- Brooks, V. B. Some examples of programmed limb movements. Brain Research, 1974, 71, 299-308.
- Easton, T. On the normal use of reflexes. American Scientist, 1972, 60, 591-599.
- Fitts, P. M. The information complexity of the human motor system in controlling the amplitude of movement. Journal of Experimental Psychology, 1954, 47, 381-391.
- Fowler, C. A., & Turvey, M. T. Skill acquisition: An event approach with special reference to searching for the optimum of a function of several variables. In G. E. Stelmach (Ed.), Information processing in motor control and learning. New York: Academic Press, 1978, 1-40.
- Grillner, S. Locomotion in vertebrates: Central mechanisms and reflex interaction. Physiological Reviews, 1975, 55, 247-304.
- Gurfinkel, V. S., Kots, Y. M., Pal'tsev, Y. I., & Fel'dman, A. G. The compensation of respiratory disturbances of the erect posture of man as an example of the organization of interarticular interaction. In I. M. Gelfand (Ed.), Models of the structural-functional organization of certain biological systems. Cambridge, Mass.: MIT Press, 1971.
- Kelso, J. A. S. Motor control mechanisms underlying human movement reproduction. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 529-543.
- Kerr, B. Evaluating task factors that influence selection and preparation for voluntary movements. In G. E. Stelmach (Ed.), Information processing in motor control and learning. New York: Academic Press, 1978, 55-69.
- Kirk, R. E. Experimental design: Procedures for the behavioral sciences. Belmont, Calif.: Brooks-Cole, 1968.
- Klapp, S. T. Reaction time analysis of programmed control. In R. Hutton

- (Ed.), Exercise and sports sciences reviews. Santa Barbara, Calif.: Journal Publishing Affiliates, 1977.
- Kozlovskaya, I. B., Atkin, A., Horvath, F. E., Thomas, J. S., & Brooks, V. B. Preprogrammed and feedback guided movements of monkeys. Behavioral Biology, 1974, 12, 243-248.
- Kulagin, A. S., & Shik, M. L. Interaction of symmetrical limbs during controlled locomotion. Biophysics, 1970, 15, 171-178.
- Luria, A. R. The working brain. New York: Basic Books, 1973.
- Orlovskii, G. N. The effect of different descending systems on flexor and extensor activity during locomotion. Brain Research, 1972, 40, 359-371.
- Schmidt, R. A. A schema theory of discrete motor skill learning. Psychological Review, 1975, 82, 225-260.
- Shik, M. L., & Orlovskii, G. N. Neurophysiology of locomotor automatism. Physiological Reviews, 1976, 56, 465-501.
- Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In G. E. Stelmach (Ed.), Information processing in motor control and learning. New York: Academic Press, 1978.
- Turvey, M. T. Preliminaries to a theory of action with reference to vision. In R. Shaw & J. Bransford (Eds.), Perceiving, acting, and knowing. Hillsdale, N.J.: Erlbaum, 1977, 211-266.
- Turvey, M. T., Shaw, R. E., & Mace, W. Issues in the theory of action. In J. Requin (Ed.), Attention and performance VII. Hillsdale, N.J.: Erlbaum, 1978.

FOOTNOTES

¹A favorite example is the difficulty an individual often encounters when attempting to rub the stomach and pat the head at the same time.

²In fact we were later to find out, via high-speed cinematographical techniques, that the hand performing the easy task did not "wait" for its more difficult counterpart, but rather moved at an entirely different velocity (see Figure 4).

³This potentially confounding problem was raised by John Morton at a preliminary presentation of the data to the Medical Research Council, Applied Psychology Unit, Cambridge, England, to whom we are grateful.

⁴The stance or support phase is the interval in the step cycle during which the foot is in contact with the ground. The swing or transfer phase refers to the period of limb retrieval for the next step.

THE CONTRIBUTION OF NATURAL DURATIONS TO SPEECH SYNTHESIZED BY FOVE RULES*

Frances Ingemann†

Abstract. Rules for speech synthesis using the FOVE program incorporate durational rules that produce differences in durations that are intended to be similar to those in natural speech. A recent set of such rules produced speech in which the words were approximately 85% intelligible when heard in moderately difficult sentences. To determine whether further attention to durational values would prove profitable in a revision of the rules, a set of experiments was conducted in which synthetic sentences by rule were modified by changing the durations to match those in a reading of the same sentences by a human speaker. Frequencies and amplitudes of the speech by rule were unchanged. Listeners' performances improved when natural durations were used, but the improvement was more noticeable at the beginning of a listening session than after subjects had had an opportunity to adapt to synthetic speech.

INTRODUCTION

Rules for speech synthesis must incorporate durational rules as well as specifications for frequency and amplitude. The FOVE program allows the user to specify these values for phonemes as well as to provide rules for changing these values in specific contexts. FOVE is essentially Kuhn's (1973) OVEBORD program, as slightly modified by Ignatius Mattingly. The program and the quality of speech produced by three sets of rules is described in Ingemann (1978). As reported there, improvements in the rules increased the intelligibility of a set of moderately difficult test sentences from 75% in 1974 to 84% in 1977 for words and from 83% to 91% for phonemes, with vowels slightly more intelligible in both versions (see Table 1).

*A version of this paper was presented at the 94th meeting of the Acoustical Society of America, Miami Beach, December 1977.

†University of Kansas.

Acknowledgment. This research was supported by Veterans Administration Contract V101 (134) P-342 and BRSG Grant RR-05596.

Table 1

Percentage Correct Responses to Intended Stimuli in 42 Sentences.

	<u>1974 Rules</u>	<u>1977 Rules</u>
Words	75%	84%
All Phonemes	83%	91%
Consonants	81%	90%
Vowels	85%	93%

Impressions gained from these earlier studies suggested that, to achieve further improvements, the rules for specifying both fundamental frequency and segment durations needed refinement. Of these two candidates, segment duration appeared to be particularly promising and was, therefore, chosen for initial examination. However, from the outset it was not clear whether durational improvements would simply add greater naturalness or whether they might contribute significantly to intelligibility. Hence, to determine whether durational improvements alone would make a difference in intelligibility, two sets of stimuli were devised in which sentences synthesized by the 1977 rules were modified so that the durations of the synthesized sentences matched those of the same sentences spoken by the investigator.

EXPERIMENT I

Listeners

The listeners were 12 students and staff members at the University of Kansas--6 for each test version. None of them had previously participated in a synthetic speech experiment.

Stimuli

Four sentences in synthetic speech (from the 17 meaningful sentences given in Table 2) were modified so that their segmental durations matched those of real speech. Durations were measured on spectrograms of real speech that had been segmented into identifiable portions: stop closure, aspiration, vocalic portions, etc. The corresponding segments in the synthetic speech were adjusted to match those of real speech by deleting or copying 5 msec time frames at equal intervals throughout the segment in order not to change the frequency and amplitude values supplied by rule. Most of the changes involved deletion of time frames since durational values for the 1977 rules had been increased slightly from earlier sets of rules in response to listener comments that the speech was too fast. (The 1977 rules produce speech at approximately 140 words per minute.)

Table 2

Sentences Used to Test the 1977 Rules and Modified for Duration.

1. Thousands of years ago migrant tribes roamed the woodlands.
- *2. The subchaser eased down the slipway.
- *3. Should backpackers eat flapjacks before starting a hike?
- *4. Right-to-work laws give employees the choice of not joining a union.
- *5. Whoever heard of a fife and drum corps wearing olive drab work clothes?
6. The math teacher and the life guard struck up a friendship in the unemployment line.
7. My clothier had the illusion that he could change any dumpy form into a svelt figure.
8. A hot dog vendor sang the national anthem.
9. Although misspelled, the embellished sign obviously got results.
10. Queen Dido wore a jeweled crown.
11. Her throne stood on a thick Persian rug.
12. Are goblins apt to live under gnarled oaks?
13. The Asian's vision of a fawn surrounded by coins was a good omen that his boy would be wealthy.
- *14. The brew contained yellow bug juice.
15. Hopefully, the treasurer inserted no fudge factor in the sinking fund.
16. The town's decision to outlaw loud toys brought shouts of joy from mothers.
17. The ragpicker's drip pan occasionally overflowed into his lunch box.

* Sentences modified for duration

The durationally-modified sentences were included as numbers 2-5 of the set of 17 sentences synthesized by rule. In one test version, only sentences 2 and 3 had natural durations. In the other version, only sentences 4 and 5 had natural durations.

Procedure

The stimuli were presented individually over earphones to the listeners who were allowed to adjust the volume to a level that they found comfortable. The following instructions were printed on the answer sheet and also played to the listeners in synthetic speech prior to the beginning of the stimuli that were to be scored:

You will hear 17 sentences. Each sentence will be preceded by a number and read twice. Please write down the complete sentence--or as much of it as you understand. You do not have to write the number. The meaning of some sentences may seem odd. Don't let this bother you. Write what you hear.

There will be a pause for you to write the sentence. Just before the next sentence, you will hear the word ready. If you have not finished writing, stop the tape. When you are ready, start the tape. Do not rewind to listen to a sentence again.

The answers were scored in two ways: words correct and phonemes correct. Correct words were counted as having all phonemes correct and all words missing from the answer sheet were counted as having all phonemes misidentified. Incorrect words were matched for the closest phonetic fit to the intended word and those phonemes that were identical were counted as correct.

Results

The results are shown in Table 3. Sentences synthesized without modification (Nos. 1, 6-17) resulted in 85% of the words correct and 89% of the phonemes correct. Sentences 2-5, synthesized completely by rule, had lower scores than the other sentences produced by rule but improved when natural durations were used. With natural durations, word recognition in sentences 2-5 improved by 7% and phoneme recognition by 5%.

Table 3

Percentage Correct Responses to 17 Sentences Synthesized with 1977 Rules.

	<u>Sentences 2-5</u>		<u>Sentences 1, 6-17</u>
	<u>Natural Durations</u>	<u>By-Rule Durations</u>	<u>By-Rule Durations</u>
Words	86%	79%	85%
Phonemes	90%	85%	89%

EXPERIMENT II

Listeners

The stimuli were presented to a different group of 12 listeners at the University of Kansas. Some had taken part in listening experiments in previous years, but none were regularly exposed to synthetic speech. Each listener heard only one test version.

Stimuli

Ten "syntactically normal" nonsense sentences taken from Nye and Gaitenby (1974) were also adjusted to match the durations of natural speech in the same way as the meaningful sentences in Experiment I had been. All sentences (see Table 4) contained four monosyllabic English words in the frame "The (Adj) (N) (V) the (N)." These sentences prevent the subjects from predicting the test

words on a semantic basis. Again, there were two test versions: In the first, sentences 1-5 had natural durations; sentences 6-10 were completely by rule. In the second version, sentences 1-5 were completely by rule and sentences 6-10 had natural durations. To determine whether these sentences were representative of the larger set of 200 sentences used by Nye and Gaitenby, 50 more sentences were synthesized entirely by rule. These sentences (Sentences 101-150 from the list of Nye and Gaitenby) were presented to the listeners after the first ten.

Table 4

Syntactically-normal Sentences Synthesized by 1977 Rules and with Durations Matched to Those of Natural Speech.

1. The full leg shut the shore.
 2. The black top ran the spring.
 3. The great car met the milk.
 4. The old corn cost the blood.
 5. The short arm sent the cow.
 6. The low walk read the hat.
 7. The sick seat grew the chain.
 8. The young voice saw the rose.
 9. The fine lip tired the earth.
 10. The large group passed the judge.
-

Procedure

The procedure followed that of Nye and Gaitenby. The following instructions appeared on the answer sheet:

You will hear 10 sentences, each played only once. You will then have 10 seconds in which to fill in the blanks for each sentence. The sentences don't make sense so just write what you hear. Before you begin to write you will hear the first five sentences so that you will know what to expect. The tape will then be rewound so that you can hear them again and write your answers.

The listeners were allowed to adjust the volume to their individual comfort level.

The printed instructions for the additional 50 sentences were as follows:

You will hear 50 synthetic speech sentences, each preceded by a number and played only once. You will then have 10 seconds in which to fill in the blanks for each sentence. The sentences don't make sense so just write what you hear.

The results were scored as in Experiment I.

Results

The results given in Table 5 show an improvement of 8% in word recognition and 3% in phoneme recognition when the natural durations of the investigator's speech were used.

Table 5

Correct Responses to Intended Stimuli for Syntactically Normal
Nonsense Sentences Synthesized with 1977 Rules.
The Ten Sentences Specified Below Were Heard First.

	<u>Ten Sentence Test</u>		<u>Fifty Sentence Test</u>
	<u>Natural Durations</u>	<u>By-Rule Durations</u>	<u>By-Rule Durations Only</u>
Words	85%	77%	82%
Phonemes	92%	89%	91%

Discussion

Examination of the results of Experiments I and II revealed that sentences compared for durational differences (sentences 2-5 of Experiment I and the first ten sentences of Experiment II) scored lower when synthesized by rule than the other sentences synthesized by rule in the same experiments. Furthermore, these other sentences synthesized entirely by rule scored only 1% lower in phoneme recognition and 1% to 3% lower in word recognition than sentences with natural durations. While it is possible that the sentences in which durational differences were compared were more difficult than the other sentences, a more likely explanation of the higher scores for the other sentences is that there was a learning effect. It is worth noting in this connection that in both experiments the sentences compared for duration always occurred at the beginning of a listening session.

To check for a learning effect in Experiment II, the words correct in the first 25 of the 50 sentences synthesized entirely by rule were compared with those correct in the last 25 sentences. The results (see Table 6) show that listeners do indeed understand more of the second half than they do of the first: The second half had 7% higher word intelligibility. Since there is no reason to believe that the second 25 sentences were inherently easier than the first 25, it seems evident that a learning effect did occur.

Table 6

Percentage Correct Responses to Intended Stimuli for Syntactically
Normal Nonsense Sentences Synthesized with 1977 Rules.

	<u>Sentences 1-25</u>	<u>Sentences 26-50</u>	<u>All 50 Sentences</u>
Words	78%	85%	82%

EXPERIMENT III

Listeners

Twelve student and staff members from the University of Kansas who had not participated in Experiment II served as listeners.

Stimuli

To see to what extent natural durations would improve the speech after the initial learning period, a third experiment was run using the same stimuli as in Experiment II but with the previously final 50 sentences now preceding the ten sentences in which natural and rule-specified durations were compared.

Procedure

The procedure was the same as in Experiment II except that the order of presentation of the two parts was reversed.

Results

Table 7 compares the results of Experiment III with those of Experiment II and clearly reveals that learning affects the scores of the sentences produced by rule. Once this learning has taken place, natural durations produce only a negligible improvement.

Table 7

Percentage of Words Correct in Syntactically-normal Nonsense Sentences
in Two Test Orders.

<u>Test Order</u>	<u>Ten Sentences</u>		<u>Fifty Sentences</u>		
	<u>Natural Durations</u>	<u>Rule Durations</u>	<u>1-25</u>	<u>26-50</u>	<u>Total</u>
Exp. II (10,50)	85%	77%	78%	85%	82%
Exp. III (50,10)	86%	83%	71%	83%	77%

CONCLUSIONS

It would seem then that slightly unnatural durations cause comprehension problems initially, but that a listener can quickly adapt and can, without special training, compensate for the unnaturalness of durations of the sort contained in the 1977 rules. For listeners who will have considerable exposure to synthetic speech, priority should probably be given to improving aspects other than duration in the rules. On the other hand, for initial acceptance and for listeners not expected to listen extensively to synthetic speech, further improvement of durational specifications is important.

REFERENCES

- Ingemann, F. Speech synthesis by rule using the FOVE program. Haskins Laboratories Status Report on Speech Research, 1978, SR-54, 165-173.
- Kuhn, G. M. A two-pass procedure for synthesis by rule. Journal of the Acoustical Society of America, 1973, 54, 339. (Abstract)
- Nye, P. W., & Gaitenby, J. H. The intelligibility of synthetic monosyllabic words in short, syntactically normal sentences. Haskins Laboratories Status Report on Speech Research, 1974, SR-37/38, 169-190.

A REVIEW OF "THE SKILLS OF THE PLODDER"^{*}

Toward a Psychology of Reading

The Proceedings of the CUNY Conferences

Arthur S. Reber and Don L. Scarborough (Eds.)

Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977 xv + 337 pp.

Reviewed by: Ignatius G. Mattingly⁺

A philosopher of science might complain that there cannot be a psychology of reading, any more than there can be a psychology of dish-washing or of bill-paying, but only intrinsically unrelated psychologies of eye-movement, of character recognition, of language, and so on, underlying the activity of reading. His argument could be corroborated by the diversity of subject matter in the eight papers collected here. Yet, as R. C. Calfee insists in his contribution, a shrewd analysis of the pitfalls of testing, it is no simple matter to study reading skills in isolation. Moreover, certain common themes recur often enough in this book to justify its title. Many of them are introduced in two long papers (L. R. Gleitman and P. Rozin; Rozin and Gleitman) really forming a comprehensive and insightful psycholinguistic treatise on "the structure and acquisition of reading" that could well have been published separately.

One such theme is the effect of orthographic structure on reading. Rozin and Gleitman make the usual point that while the principle of a logographic system is easier to grasp, a phonographic (syllabary or alphabetic) system, once understood, facilitates analysis of unfamiliar words. But a logographic system and a phonographic system each have a further distinct advantage lacking in the other, as L. Brooks shows, in what is certainly the most original paper in the book. In experiments with artificial character sets, he finds that, even if only six different words are to be remembered, an alphabetic four-character representation of a word, once learned, is read faster than an arbitrary four-character representation. Even if as many as 120 different words are to be remembered, a "glyphic" representation, in which the four characters are stacked and superimposed to form a complex, visually distinct symbol, whether alphabetic or arbitrary, is read faster than a representation in which the four characters appear in horizontal sequence. Brooks' results support J. Williams' observation, in her perceptive account of her work with the learning-disabled, that the "whole word" method is not a desirable strategy for teaching children to read an alphabetic orthography. They also imply that, in principle, the advantages of phonological correspondence and visual distinctiveness could be combined in an orthography that was

^{*} A revised version of this review appeared in Contemporary Psychology, 1978, 23, 731-732.

⁺ Also, University of Connecticut, Storrs.

Acknowledgment: This review was written with support from NICHD Grant HD-01994.

both phonographic and glyphic. It is interesting that before the advent of printing, alphabetic scripts made more common use of devices that are moves in the glyphic direction, like the tilde over a letter to represent following n; and that there are no actual writing systems that are neither phonographic nor glyphic.

A familiar controversy provides another theme: Is the reader a "plodder" (to borrow Rozin and Gleitman's terms, p. 59) who proceeds letter by letter, or an "explorer," who samples the printed page selectively to confirm his educated guesses? Rozin and Gleitman themselves believe that the truth lies somewhere in between. These two hypotheses, however, are usually formulated by their proponents so vaguely as to raise a doubt whether they can serve as endpoints of a meaningful continuum. But for what it is worth, the evidence in other studies reported here is all on the side of the plodder. K. Rayner and G. W. McConkie have ingeniously experimented with a computer-controlled system that can track a reader's eye movements and modify the text on a CRT display as he reads it. Their subjects (reading textbook material, to be sure, and anticipating a comprehension test) progress quite methodically from left to right, have a surprisingly narrow "perceptual span" within which they can identify words during a fixation, and tend to fixate longer on more difficult words. And W. Kintsch, studying the semantic structure of texts, finds that reading time for a text is quite sensitive to the number of elementary propositions and the number of distinct propositional arguments in the text base. Neither of these findings offers much encouragement for the "explorer" hypothesis.

The special kind of awareness that a child must develop in order to read an alphabetic orthography is stressed by several contributors. But there seem to be various misunderstandings about what the child can be and must become aware of. Exercises in blending and segmentation serve to awaken the child's linguistic institutions, but Williams (along with many other students of reading) calls these skills "auditory," (pp. 283-285). I. Y. Liberman, D. Shankweiler, A. M. Liberman, C. Fowler, and F. W. Fischer, who give an illuminating account of the performance on certain linguistic tasks of good and poor readers, understand about linguistic awareness very well, yet they suggest that the relative inaccessibility of linguistic units depends on the degree to which they are encoded in the speech signal: Their subjects are said to count syllables more accurately than phonemes because the former are less encoded than the latter (p. 210). But if a child counts syllables accurately, it is because he has access, not to encoded acoustic information, but to representations of phonological syllables in his mental lexicon (e.g., for an utterance such as [sku], speakers of English and of Japanese would give different, but equally, correct responses). Such access is probably facilitated by the phonological (not phonetic or acoustic) identity between one-syllable words and the component syllables of longer words. Rozin and Gleitman, going a bit further, argue that learning to read "requires...gaining access to the machinery in the head which analyzes and produces sound segments" (p. 56). But gaining access to highly encoded segments through the machinery of speech perception is probably impossible and surely unnecessary. The child's task is rather to relate orthographic representations just to the output of the perceptual and linguistic machinery: phonological representations. Access to phonological segments has to be achieved by analysis of the larger phonological units of which the child is already aware: syllables and

words. The encodedness of speech is relevant to linguistic awareness only in that it underlies a pedagogical difficulty: Since encoded sounds cannot readily be uttered in isolation, the teacher cannot refer to the phoneme /b/ by saying "[b]," but if he says, "[b]," he may mislead the student.

Much more might be said about these papers, every one of which is lucid, thoughtful and in one way or another provocative. The editors have done a service in making them available.

II. PUBLICATIONS

III. APPENDIX

PUBLICATIONS

- Abramson, A. S. Lexical tone and sentence prosody in Thai. In Proceedings of the Ninth International Congress of Phonetic Sciences. Copenhagen: University of Copenhagen, 1979, 380-387.
- Laer, T. Single-motor unit electromyographic analysis of fundamental frequency control during phonation. In C. J. Deluca (Ed.), Proceedings of the 4th Congress of the International Society of Electrophysiological Kinesiology. Boston: International Society of Electrophysiological Kinesiology, 1979, 208-209.
- Baer, T., Bell-Berti, F., & Rubin, P. Articulation and voice quality. In V. Lawrence (Ed.), Transcripts of the Seventh Symposium on Care of the Professional Voice. New York: The Voice Foundation, 1979, 48-53.
- Bell-Berti, F. Inter-articulator programming and clinical advice to "slow down." Journal of Childhood Communication Disorders, in press.
- Bell-Berti, F., Baer T., Harris, K. S., & Niimi, S. Coarticulatory effects of vowel quality on velar elevation. Phonetica, 1979, 36, 187-193.
- Bell-Berti, F., & Harris, K. S. Anticipatory articulation: Some implications from a study of liprounding. Journal of the Acoustical Society of America, 1979, 65, 1268-1270.
- Bell-Berti, F., Raphael, L. J., Pisoni, D. B., & Busch, J. R. Some relationships between speech production and perception. Phonetica, in press.
- Borden, G. J. An interpretation of research on feedback interruption in speech. Brain and Language, 1979, 7, 307-319.
- Borden, G. J., & Gay, T. Temporal aspects of articulatory movements for /s/-stop clusters. Phonetica, 1979, 36, 21-31.
- Harris, K. S., & McGarr, N. S. Relationships between speech perception and speech production in normal hearing and hearing impaired speakers: Implications for training. In Proceedings of the Institute on Special Assessment and Speech Improvement in the Hearing Impaired, National Technical Institute for the Deaf, Rochester, N.Y., 22 June 1979. Washington, D.C.: Alexander Graham Bell Association, in press.
- Kitzing, P., & Löfqvist, A. Evaluation of voice therapy by means of photoglottography. Folia Phoniatrica, 1979, 31, 103-109.
- McGarr, N. S. Evaluation of speech in intermediate school-aged deaf children. In Proceedings of the Institute on Special Assessment and Speech Improvement in the Hearing Impaired, National Technical Institute for the Deaf, Rochester, N.Y., 22 June 1979. Washington, D.C.: Alexander Graham Bell Association, in press.
- Repp, B. H. Stimulus dominance in fused dichotic syllables: Trouble for the category goodness hypothesis. Journal of the Acoustical Society of America, in press.

APPENDIX

DDC (Defense Documentation Center) and ERIC (Educational Resources Information Center) numbers SR-21/22 to SR-58:

<u>Status Report</u>	<u>DDC</u>	<u>ERIC</u>
SR-21/22	AD 719382	ED-044-679
SR-23	AD 723586	ED-052-654
SR-24	AD 727616	ED-052-653
SR-25/26	AD 730013	ED-056-560
SR-27	AD 749339	ED-071-533
SR-28	AD 742140	ED-061-837
SR-29/30	AD 750001	ED-071-484
SR-31/32	AD 757954	ED-077-285
SR-33	AD 762373	ED-081-263
SR-34	AD 766178	ED-081-295
SR-35/36	AD 774799	ED-094-444
SR-37/38	AD 783548	ED-094-445
SR-39/40	AD A007342	ED-102-633
SR-41	AD A013325	ED-109-722
SR-42/43	AD A018369	ED-117-770
SR-44	AD A023059	ED-119-273
SR-45/46	AD A026196	ED-123-678
SR-47	AD A031789	ED-128-870
SR-48	AD A036735	ED-135-028
SR-49	AD A041460	ED-141-864
SR-50	AD A044820	ED-144-138
SR-51/52	AD A049215	ED-147-892
SR-53	AD A055853	ED-155-760
SR-54	AD A067070	ED-161-096
SR-55/56	AD A065575	ED-166-757
SR-57	**	ED-170-823
SR-58	**	**

AD numbers may be ordered from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22151

ED numbers may be ordered from:

ERIC Document Reproduction Service
Computer Microfilm International
Corp. (CMIC)
P.O. Box 190
Arlington, Virginia 22210

Haskins Laboratories Status Report on Speech Research is abstracted in Language and Behavior Abstracts, P.O. Box 22206, San Diego, California 92122.

**DDC and/or ERIC order numbers not yet assigned.